

RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite *Plasmodium falciparum*."
RT *Nature* 419:458-511(2002).
RL EMBL; AB014841; AAN5938.1; -.
DR HSSP; O93JC2; IMWT.
KW Hypothetical protein.
SQ SEQUENCE 2227 AA; 262841 MW; 6D5D4B8EF43339C CRC64;

This Page Blank (uspto)

LENGTH: 1787
TYPE: PRT
ORGANISM: Kl Parasite Clone
US-10-415-253-2

Query March 5.0%; Score 200.5; DB 15; Length 1787;
Best Local Similarity 20.5%; Pred. No. 0.015;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AES--KERN-LGNGEGLKXDT-----TGVEHHQENB-----ESIX 40
DB 868 AESVTTSNILBEIQENTITNDTIEKLEBLHENVLSAALEVQSEBEKEVLDVIREVK 927
QY 41 EKSSFTI-----DRNISTIRD-FEN-----KDLKGLKKKREYDDPTS 78
DB 928 EEWATTLLETVEQAEBKSNITTELFENLENNAVESNENNAENLEKLBETVFNVLDKVE 987
QY 79 ET-----GKRMK--EYDYKY-----DDKG--NIIAYDDGTDLEYEY-----EKLD-- 114
DB 988 ETVEISGSLLENNDKAFPFSEIFPNVKGIOENLIT---GMFRSLETSLVIOSEKEDLN 1044
QY 115 -EIKSKIYGLSPSKGHPILKGISNYSKNAKV-----YGNMYKSIIEIKATIDPH 166
DB 1045 ENVVSSIIDNIENMKEG---LLNKLKLENIISTEGVQETVEHVEQVYVDVDPAMK--- 1097
QY 167 SKMTTFDIYANINDVQGLAFAGDMRLFYKNDQKKAERIKIMPEKIKETSEYVYSSY 226
DB 1098 ---DQFLGLINAGGL---KEMFPNLEDFKSSDVTVEBIEKDEPVQKEVERET 1146
QY 227 GNVIELGEGD---LSKRPDLNLTGM-----ESGKIYSDSEKQOYLKONITLRGYALX 277
DB 1147 VSIIIEEMENIVDVLEEKEDDTQMDAVESIEISSDSKETESIKOX---EKDVSIV 1203
QY 278 VTTNPG-----KTDMLEGNGVYSK--EDIAKIQCANPL----- 310
DB 1204 VEEVQDNMDESVEKYLEKMBELMKDAVEINDITSKLIERQBLNEVEADLIRDMEX 1263
QY 311 -----RALSETTIVADSRNVEDGR--STOSVLSALDGFNIIRYQVFPKMDKGEALDK 363
DB 1264 LKELEKALSE---DSKEIIDAQDITLKEVIEEHD-----ITTTLDVAVELKDV 1309
QY 364 DGNLVTSSKLVLPKGDKPEYTGBDKFVNEALIKEDGSMFLIDTKPVL-----SM 413
DB 1310 BBDKIEKXSDL---KOLBEDILKE---YKEIKELSEILIEDYKELKTIEFDILEKKEI 1362
QY 414 DKATFPNPKSKNIYRNPFTLRGKISDXGCPNMEIRVNESVVD-----NYLIYGL 465
DB 1363 EKDHKEKEEBAEBEIKDLEADLKEVS---SLEVEBEKLEEVHELKEVEYHIIISGDA 1417
QY 466 HINDTRDINIKLNVKD--GDIM-----DMGMKDYKANGPPDKVTMDGNVYLQTVSDLN 518
DB 1418 HIKGLBEDDLR-EVDDLGSLIDMLKGMELGDMKESLIEDVTTKGERV---ESLKDVL 1473
QY 519 AKAVGVHYQFLYDN-----VXPEVINIDPKGNTSIEYADKSVENINDK--RN 564
DB 1474 SSALGMDEEOKMTRKKAQRKLEBVLKBEVKBPPKKIT-----KKKVRFDIKDEKPRD 1528
QY 565 NGFDGEIQGSHIYNGK-----EYTSNDIKQIIDKTILN-----IKIYVQDFAR--N 609
DB 1529 EIVEYEMKODEIEEVEDEIDEDIDEDIDEDIDEDIDEDIDEDIDEDIDEDIDEDIDED 1588
QY 610 TTVKEPFIANKDTGE--VSELKPH-----RVTVTIONGKEMSSSTIVSEB 651
DB 1589 VRKKKKLEKRYEBEGVGLKGVADVEMKIVQKIDKEVDKESKALESKNDVTNNVLKOND 1648
QY 652 FILPYKGELEKGYQPDGWEISGFEGKCDAGYVNL-----SKDTFIKPVK 698
DB 1649 FFSKY-KAFVVKYKFAAPFISAVAA--FASYVVGFFTFSLSSCWTIASSTYLLSKVDK 1705
QY 699 KIEKKESBENR---TFDV-----SKKDNPOVNSQLNESHKEDLQR 739
DB 1706 TINKKE---RPFYSFVFDIFPNLKHLYLOQMKESKEKNNNVIEVT--NKAERKANVOY 1760

QY 740 EHSQKSDSTKVATFVLDKNNISSKSTNNPNK 773
DB 1761 TINKET-----TKVDKNNKVPKRRRTQSK 1766

Search completed: February 16, 2005, 19:32:38
Job time : 147 secs

This Page Blank (uspto)

RESULT 15
US-09-949-016-11433
Sequence 11433, Application US/09949016
Patent No. 681239
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11433
LENGTH: 2733
TYPE: PR
ORGANISM: Human
US-09-949-016-11433

Query Match 4.6%; Score 185; DB 4; Length 2733;
Best Local Similarity 20.0%; Pred. No. 0.0032;
Matches 170; Conservative 150; Mismatches 338; Indels 194; Gaps 39;

21 LKDDTGVHHHQNEN-SIKKSSPTIDRNISTIRDFENKDLKGLIKK-FREYVD 75
1319 LKERIAGLEBEKQNKESFQTLNENKYLISQIST-KDGRKMLQEBVTKNMLNQIOE 1377
76 FTSEGRMBEYDYKVD-----KGNIIAY-DGGTDLEYTEKDLBTKSKYIG 122
1378 ELSRVTLKETAEEKQDLERLAKQALANGLSIGNYCQVDYDAQIKNELLESKMLKX 1437
123 VLSPKDHPEILCKISNVSNAKAVY-----GN-----NYKSIEXA 160
1438 CVSELEBEKQQLVAKTKVSEIRKREYLEKIQAGKBERKNSHAKELQELKEKQOEVNQ 1497
161 TKVD---FHSKTMFPDLXANINDIVDGLAFAGDKMLFYKNDQKKAEI-KIRMPKIKET 216
1498 LQKDCIRYQEKISALERTVKALE-----FVQTESQKOLBITTKENLAQAVEHR 1544
217 KSEPYVSYGVNVELEBGDLSKNKPDMLTWESGKIYSDSEKQOYLKDNIIILK----- 272
1545 KKAQAEIASFVVLDDTQSEARVADNLKELQSKESYKSGMKQKQKQEDLERLEQA 1604
273 -GYALKVTTYNPGTIDMLEGNGVYSKBDIACIQRANPNLALSTTIYADSRNVEDGERST 331
1605 EEKLEKERNKQKQKDLALRRKVLERTIGRIQ-----VTLNKKQKEVQO----- 1649
332 QSVLMSALDGNITRIYQVFTFRMNDKGEALIDKGLVTDSSKLVLPFGKDDKXYTGEDKFN 391
1650 ---LQENLDS-TVQTLAAFTSMS---SLQDDRDRTVDEAK---KMERKPSDAIQSK 1696
392 VEAI--KEDG-SMLFIDTKPVNLMDKNYFNPSSK---SNKIY--VRNPEFIARGKISDK- 442
1697 EEBIRLEEDNCSYKQDLQKQSIMBELKINISRLNHHKQIMBSKAQTEVOLQKXVCDTL 1756
443 GGFNWEELRVNESVVDNLYIGDLHIIDTRFNKILANVKQGDIMQGMKDYKANGFPDRTY 502
1757 QGENKEILL-----SOLBRTR-----HLTHSSONELAKLESBIKSLKQDLT 1796
503 DMD-----GNV--YLQTSYSDL-NAKAVGVHYQFLYDNVYKPEV----- 537
1797 DLSNLSKCKEQKQKALBGITIQOEADIQNSK-----FSTYQULFTDLOASRELTSLAH 1849
538 NIDPKGNTSIEYADGK--SVVFNINDKRNNGFDSIEQOHYIYNGKEYTSF---NDIKQI 592
1850 EINNKEQKISILSGKEBALQVALAELRQO-HDKKIKELNMLLSQEBEENIVLEBENKKA 1908

QY 593 IDKTLNIKIVKDPARTTVAIFILNKDQGEVSBK--PHRVTVTIQNGKMSSTIVSBE 650
DB 1909 VDKTNQMLBETIKTKENIQKQALDSFVKSMSLQNDRDIIVGDYQQLBRRHSTILEK 1968
QY 651 DPILPVTKGELKGYQPDGWEISGFEQ-----KQDAGVIVILSKDT 691
DB 1969 DQLIQBAABENNKUKE---SIRGLRSHMDILNSENAKDLABLQYREDLNVITTI-KDS 2023
QY 692 FIKQVEK-KIEKKKBEENKPTFVYSKKQNDPQVNHSSQNES-----HKKEDLQREBSQK 745
DB 2024 QQKQULBVLQQNNBLENKTKAKLEBKLESSEAN-EDLRSPNALQEBKQDLSKELISBK 2082
QY 746 ---SDSTEDVTA 754
DB 2083 VSIQULTRQVTA 2094

Search completed: February 16, 2005, 19:20:41
Job time : 50 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:06:53 ; Search time 170 Seconds
(without alignments)
1758.622 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGELAEKFKMLGNKGEGS.....ATVLDRNLSKSTTNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_16Dec04:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003as:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	4 AAB48343	Aab48343 S. pneumo
2	4026	100.0	2140	6 ABU01020	Abu01020 S. pneumo
3	4026	100.0	2140	6 ABU45746	Abu45746 Protein e
4	4026	100.0	2140	8 ADM82113	Adm82113 S. pneumo
5	4015.5	99.7	2120	3 AAY81710	Aay81710 Streptoco
6	3789.5	94.1	2138	8 AdK48759	AdK48759 Streptoco
7	3048	75.7	637	8 ADR94534	Adr94534 Novel S.
8	750.5	18.6	1529	8 ADR86136	Adr86136 Novel S.
9	615	15.3	117	2 AAW5096	Aaw5096 Streptoco
10	615	15.3	117	5 ABP54590	Abp54590 S. pneumo
11	615	15.3	117	7 ADC45149	Adc45149 S. pneumo
12	227	5.6	1196	6 ABU24813	Abu24813 Protein e
13	222	5.5	861	7 ABO23608	AbO23608 Plasmodu
14	220	5.5	1166	8 ADK9186	AdK9186 Streptoco
15	218	5.4	1639	2 AAW54145	Aaw54145 P. faicip
16	218	5.4	1639	5 AAE29345	Aae29345 Plasmodu
17	217	5.4	1166	8 AdK9185	AdK9185 Streptoco
18	217	5.4	1166	8 AdK9178	AdK9178 Streptoco
19	217	5.4	1166	8 AdK9184	AdK9184 Streptoco
20	216	5.4	1166	8 AdK9183	AdK9183 Streptoco
21	216	5.4	1233	5 ABP30203	Abp30203 Streptoco
22	216	5.4	1233	5 ABP29675	Abp29675 Streptoco
23	216	5.4	1233	8 AdK9176	AdK9176 Streptoco
24	216	5.4	1233	8 ADO10469	Ado10469 Group B S
25	216	5.4	1239	5 ABP25822	Abp25822 Streptoco

26	215.5	5.4	1254	2 AAR07503	Aar07503 Merozite
27	215.5	5.4	1254	2 AAW24575	Aaw24575 Merozite
28	215	5.3	1166	8 AdK9179	AdK9179 Streptoco
29	212.5	5.3	2485	3 AAB18172	Aab18172 Plasmodu
30	212	5.3	1141	6 ABU42327	Abu42327 Protein e
31	212	5.3	1188	3 AAB18183	Aab18183 Plasmodu
32	211.5	5.3	991	8 ADP86452	Adp86452 Clostridi
33	211.5	5.3	1104	7 ADG73658	Adg73658 C. perfri
34	209	5.2	1558	3 AAB18324	Aab18324 Plasmodu
35	208.5	5.2	3029	8 ADP25433	Adp25433 Plasmodu
36	207.5	5.2	4688	6 ABU48941	Abu48941 Protein e
37	207	5.1	1191	6 ABU24124	Abu24124 Protein e
38	207	5.1	1654	1 AAP50777	Aap50777 Sequence
39	207	5.1	1979	3 AAB18171	Aab18171 Plasmodu
40	205	5.1	2024	8 ADP25444	Adp25444 Plasmodu
41	204.5	5.1	1516	3 AAB18195	Aab18195 Plasmodu
42	204.5	5.1	2500	3 AAB18272	Aab18272 Plasmodu
43	204	5.1	1166	6 ABJ18982	Abj18982 Pathogen
44	204	5.1	1245	3 AAB18244	Aab18244 Plasmodu
45	203	5.0	1166	2 AAY08643	Aay08643 S. aureus

ALIGNMENTS

RESULT 1	
ID	AAB48343 standard; protein; 773 AA.
AC	AAB48343;
DT	20-APR-2001 (first entry)
XX	
DE	S. pneumoniae Sp130 polypeptide.
XX	
KW	Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
KW	bronchial; lung; blood; infection; immune response; immunotherapy;
KW	antibacterial; auditory; vaccine.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	WO200076540-A2.
PD	21-DEC-2000.
XX	
PF	09-JUN-2000; 2000MO-US015925.
XX	
PR	10-JUN-1999; 99US-0138453P.
XX	
PA	(MEDI-) MED IMMUNE INC.
XX	
PI	Adamou JB, Choi GH;
XX	
DR	WPI; 2001-112197/12.
DR	N-PSDB; AAC84742.
XX	
PT	New vaccines comprising Sp128 or Sp130 polypeptides, for treating and
PT	preventing pneumococcal infections, particularly infections caused by
PT	Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
PT	blood infections.
XX	
PS	Claim 8; Page 51-54; 54pp; English.
XX	
CC	The invention relates to novel immunogenic polypeptides, Sp128 and Sp130
CC	from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC	the treatment and prevention of pneumococcal infections, particularly
CC	infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC	bronchial, lung or blood infections. The antigens are used as immunogeni
CC	c agents to stimulate an immune response. The antisera and antibodies may
CC	also be used in diagnosing and treating pneumococcal infections.
CC	Recombinant polypeptides serve as a mechanism for stimulating production
CC	of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC	as reagents in other processes such as affinity chromatography. The

CC Present sequence represents the *S. pneumoniae* Sp130 polypeptide
XX Sequence 773 AA;
SQ
Query Match 100.0%; Score 4026; DB 4; Length 773;
Best Local Similarity 100.0%; Pred. No. 7, 5e-225;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESKSTIDRNISTIRDPENK 60
DB 1 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESKSTIDRNISTIRDPENK 60
QY 61 DLKKLKKKKFREVDFTSETGKMEEDYKYDDKNNIIAYDGDGLDEYETKLDKISKXI 120
DB 61 DLKKLKKKKFREVDFTSETGKMEEDYKYDDKNNIIAYDGDGLDEYETKLDKISKXI 120
QY 121 YGVLSPSKQGHFILLKISNVSKNAAYYGNKYKSEIKATKIDFSKMTFPLYNINID 180
DB 121 YGVLSPSKQGHFILLKISNVSKNAAYYGNKYKSEIKATKIDFSKMTFPLYNINID 180
QY 181 IYDGLAFAGDMRLFVKNDQKAEIKIRMEPIKETSEYVSSYGVNIELEGDLSKN 240
DB 181 IYDGLAFAGDMRLFVKNDQKAEIKIRMEPIKETSEYVSSYGVNIELEGDLSKN 240
QY 241 KPDNLTKMSSGKTYSDSEKQOYLKNNIIRKGYALKATTYNGKTDMLBGNVSKED1 300
DB 241 KPDNLTKMSSGKTYSDSEKQOYLKNNIIRKGYALKATTYNGKTDMLBGNVSKED1 300
QY 301 AKIQKANPMLRALSETTIYADSHNVEDGRSTOSVLSALDGFNIIYYOFTFPMNDKGA 360
DB 301 AKIQKANPMLRALSETTIYADSHNVEDGRSTOSVLSALDGFNIIYYOFTFPMNDKGA 360
QY 361 IDKQGNLVTDSSKLVLFQKODKEYTGEDKFNVAIKEDGSMLEFIDTKPVNLSMDKNYFNP 420
DB 361 IDKQGNLVTDSSKLVLFQKODKEYTGEDKFNVAIKEDGSMLEFIDTKPVNLSMDKNYFNP 420
QY 421 SSSNKTIVNRPETLYLFGKIDSGKGFNWEKRVESVVDNVLTYGDLHIDNTRDNLANK 480
DB 421 SSSNKTIVNRPETLYLFGKIDSGKGFNWEKRVESVVDNVLTYGDLHIDNTRDNLANK 480
QY 481 DGDIMDMGKMDYKANGFPDKVTMDGNVYLQTSYSDLNKAVGVHOPYLVNKPPEVND 540
DB 481 DGDIMDMGKMDYKANGFPDKVTMDGNVYLQTSYSDLNKAVGVHOPYLVNKPPEVND 540
QY 541 PKGNISIEYADGKSVFNNDRKNNPDEBIOEHYIYNGKEYTSFNDIKQIIDKTLN1K 600
DB 541 PKGNISIEYADGKSVFNNDRKNNPDEBIOEHYIYNGKEYTSFNDIKQIIDKTLN1K 600
QY 601 IYVQDPAARTTYKEFLINKDTGVSSELKPHRVTVTIQNGKEMSTTVSEBDFLLPYKGE 660
DB 601 IYVQDPAARTTYKEFLINKDTGVSSELKPHRVTVTIQNGKEMSTTVSEBDFLLPYKGE 660
QY 661 LEKGYQFDGMEISGFEGKDGAGVNLISKDTPIKPVFKKLEKEEENKPTPDVSKKON 720
DB 661 LEKGYQFDGMEISGFEGKDGAGVNLISKDTPIKPVFKKLEKEEENKPTPDVSKKON 720
QY 721 POWNHSQNLNESHKEDLQREBHSQKSDSTKDVATVATVLDKNNISSKSTNNPNK 773
DB 721 POWNHSQNLNESHKEDLQREBHSQKSDSTKDVATVATVLDKNNISSKSTNNPNK 773

RESULT 2
ABU01020
ID ABU01020 standard; protein; 2140 AA.
XX
XX
AC ABU01020;
XX
XX
DT 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
DE *S. pneumoniae* type 4 strain protein from coding region #590.
XX
XX
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;

KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
XX OS Streptococcus pneumoniae; type 4 strain.
XX
XX PN WO200277021-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 27-MAR-2002; 2002WO-1B002163.
XX
XX PR 27-MAR-2001; 2001GB-00007658.
XX
XX PA (CHIR-) CHIRON SPA.
XX
XX PA (GENO-) INST GENOMIC RES.
PI Masignani V, Tettelin H, Fraser C;
XX
XX DR WPI, 2003-040579/03.
XX
XX DR N-PSDB; ABX06302.
XX
XX PT New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
XX
XX PT useful as medicaments for treating or preventing a disease or infection
XX
XX PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX
XX PT ear infection.
XX
XX PS Claim 1; SEQ ID NO 1180; 56pp; English.
XX
XX CC The invention relates to a protein comprising or having at least 50%
XX
XX CC identity to any of the 2469 amino acid sequences, identified in the
XX
XX CC specification (available on a computer readable format), or its fragment,
XX
XX CC expressed from 2469 of 2489 identified DNA coding regions from the
XX
XX CC *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as
XX
XX CC AB556454. Also included are an antibody which binds one of the proteins,
XX
XX CC treating a patient by administering the protein, DNA or antibody (in a
XX
XX CC composition), a kit comprising first and second primers, which are the
XX
XX CC nucleic acid ciled above or fragments between nucleotides 8-100 of a
XX
XX CC sequence not defined in the specification, for amplifying a target
XX
XX CC sequence contained within a *Streptococcus* nucleic acid sequence, where
XX
XX CC the first primer is substantially complementary to the target sequence
XX
XX CC and the second primer is substantially complementary to the complement of
XX
XX CC the target sequence, and where the parts of the primers having
XX
XX CC substantial complementarity define the termini of the target sequence to
XX
XX CC be amplified, assay comprising contacting a test compound with the
XX
XX CC protein, and determining whether the test compound binds to the protein
XX
XX CC and a *Streptococcus pneumoniae* bacterium, where one or more genes
XX
XX CC encoding the proteins has been rendered inactive. The proteins, nucleic
XX
XX CC acid molecules, antibody and compositions are useful as medicaments for
XX
XX CC treating or preventing a disease or infection due to streptococcus
XX
XX CC bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis
XX
XX CC media or ear infection. They are also useful in developing vaccines,
XX
XX CC diagnostics and antibiotics. The methods are useful for identifying
XX
XX CC immunodominant proteins. The present sequence is one of the 2469 proteins
XX
XX CC expressed by the identified coding regions from the genomic sequence.
XX
XX CC Note: The sequence data for this patent did not form part of the printed
XX
XX CC specification, but was obtained in electronic format directly from WIPO
XX
XX CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
XX
XX CC standardise OS field)
SQ Sequence 2140 AA;
Query Match 100.0%; Score 4026; DB 6; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2, 9e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESKSTIDRNISTIRDPENK 60
DB 1334 KLGELAESKPKNLGNGKESGLKKDTTGVHHQENESIKESKSTIDRNISTIRDPENK 1393
QY 61 DLKKLKKKKFREVDFTSETGKMEEDYKYDDKNNIIAYDGDGLDEYETKLDKISKXI 120
DB 1394 DLKKLKKKKFREVDFTSETGKMEEDYKYDDKNNIIAYDGDGLDEYETKLDKISKXI 1453

QY 121 YGVLPSPKDGHEFLIKISNVSKNAKYVGNVSKYIEIKATKYDFHSTMTFPLVANIIND 180
DB 1454 YGVLPSPKDGHEFLIKISNVSKNAKYVGNVSKYIEIKATKYDFHSTMTFPLVANIIND 1513
QY 181 YVDGLAFAGDMRLFYVNDNQKAEIKIRMEPEIKETKSEYPIVSSYGVNIEIGEGLSLKN 240
DB 1514 YVDGLAFAGDMRLFYVNDNQKAEIKIRMEPEIKETKSEYPIVSSYGVNIEIGEGLSLKN 1573
QY 241 KPDNLTKMESGKIYSPSEKQVLLKNNILIRKGYALKVTTYNPKTMDLEGNVSKEDI 300
DB 1574 KPDNLTKMESGKIYSPSEKQVLLKNNILIRKGYALKVTTYNPKTMDLEGNVSKEDI 1633
QY 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFKMDKGBA 360
DB 1634 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFKMDKGBA 1693
QY 361 IDKQGNLYVDSKLVFGKDDKKEYTGDKFNVAIEKEDGSMFLIDTRPVNLSMDKNYFNP 420
DB 1694 IDKQGNLYVDSKLVFGKDDKKEYTGDKFNVAIEKEDGSMFLIDTRPVNLSMDKNYFNP 1753
QY 421 SKSNKIYVNPPEFYLGKISDKGPFNMLRVNNSVVDNYLYIGDLHIDNTRDPRNKLANKY 480
DB 1754 SKSNKIYVNPPEFYLGKISDKGPFNMLRVNNSVVDNYLYIGDLHIDNTRDPRNKLANKY 1813
QY 481 DGDIMDMGKDYKANGFPDKVYTDMDGNVYLGQTSYSDLNAAKAVGHYQFLVDNYPKPEVNI 540
DB 1814 DGDIMDMGKDYKANGFPDKVYTDMDGNVYLGQTSYSDLNAAKAVGHYQFLVDNYPKPEVNI 1873
QY 541 PKGNISIEYADGKSVFNNIDKRNNGPDGEIOGHYIYNGKEYTSFNDIKQIIDKTLNLIK 600
DB 1874 PKGNISIEYADGKSVFNNIDKRNNGPDGEIOGHYIYNGKEYTSFNDIKQIIDKTLNLIK 1933
QY 601 IYVKDARATTVYEFILNKDTGEVSELKPHRVVTIIONGKEMSTIYSEDFILPYKGE 660
DB 1934 IYVKDARATTVYEFILNKDTGEVSELKPHRVVTIIONGKEMSTIYSEDFILPYKGE 1993
QY 661 LEKGYOPDMEISGFEKGDAGVIVNLKDTFIKPYFKKIEKKKEENKTFPDVSKKON 720
DB 1994 LEKGYOPDMEISGFEKGDAGVIVNLKDTFIKPYFKKIEKKKEENKTFPDVSKKON 2053
QY 721 PQVNHQALMESHRKEDLOREHSHQKSDSTQVATATYLDKNNISSKSTNNPNK 773
DB 2054 PQVNHQALMESHRKEDLOREHSHQKSDSTQVATATYLDKNNISSKSTNNPNK 2106

RESULT 3
ABU45746
ID ABU45746 standard; protein; 2140 AA.
AC ABU45746;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #31273.
XX
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JM,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX WPI; 2003-029926/02.
XX N-PSDB; ACA49616.
DR
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 73670; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 2140 AA;
Query Match 100.0%; Score 4026; DB 6; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.9e-224;
Matches 773; Mismatches 0; Indels 0; Gaps 0;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLGELAESKFKNLGNGEGSLKKDITGVHHOENESIKESFTIDNISTIRPFENK 60
DB 1334 KLGELAESKFKNLGNGEGSLKKDITGVHHOENESIKESFTIDNISTIRPFENK 1393
QY 61 DLKLLIKKKFREVDFTSETGRMEBYDYKDDKNNIAYVDDGTDLLEYETKLDLISXI 120
DB 1394 DLKLLIKKKFREVDFTSETGRMEBYDYKDDKNNIAYVDDGTDLLEYETKLDLISXI 1453
QY 121 YGVLPSPKDGHEFLIKISNVSKNAKYVGNVSKYIEIKATKYDFHSTMTFPLVANIIND 180
DB 1454 YGVLPSPKDGHEFLIKISNVSKNAKYVGNVSKYIEIKATKYDFHSTMTFPLVANIIND 1513
QY 181 YVDGLAFAGDMRLFYVNDNQKAEIKIRMEPEIKETKSEYPIVSSYGVNIEIGEGLSLKN 240
DB 1514 YVDGLAFAGDMRLFYVNDNQKAEIKIRMEPEIKETKSEYPIVSSYGVNIEIGEGLSLKN 1573
QY 241 KPDNLTKMESGKIYSPSEKQVLLKNNILIRKGYALKVTTYNPKTMDLEGNVSKEDI 300
DB 1574 KPDNLTKMESGKIYSPSEKQVLLKNNILIRKGYALKVTTYNPKTMDLEGNVSKEDI 1633
QY 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFKMDKGBA 360
DB 1634 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFKMDKGBA 1693

KW pneumococcal septicemia; sinusitis; meningitis; therapy.
 XX Streptococcus pneumoniae.
 XX MO200006738-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB002452.
 XX
 PR 27-JUL-1998; 98GB-00016336.
 PR 19-MAR-1999; 99US-0123529P.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
 XX WPI, 2000-195301/17.
 DR N-PSDB; AA291806.
 XX
 PT Streptococcal proteins and polynucleotides useful for diagnosis,
 PT treatment and prophylaxis of bacterial infections.
 XX
 XX Claim 2; Page 41-42; 76pp; English.
 PS
 SS This sequence represents a Streptococcus pneumoniae protein of the
 CC invention. The proteins (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae
 CC infection. As the sequences can be used to treat S. pneumoniae infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicemia, otitis media, sinusitis, and
 CC meningitis
 CC
 CC
 SQ Sequence 2120 AA;
 Query Match 99.7%; Score 4015.5; DB 3; Length 2120;
 Best Local Similarity 99.9%; Pred. No. 1.1e-223;
 Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 KLGEIAESKRNKNGKEGSLKKDTTGVHHQENESIKESPTIDRNISTIRDFENK 60
 DB 1313 KLGEIAESKRNKNGKEGSLKKDTTGVHHQENESIKESPTIDRNISTIRDFENK 1372
 QY 61 DLKLLIKKKFREYDDTSETGKMEYDYKYDKGNIIAYDGTDLLEYETKDEIKSKI 120
 DB 1373 DLKLLIKKKFREYDDTSETGKMEYDYKYDKGNIIAYDGTDLLEYETKDEIKSKI 1432
 QY 121 YGVLSPSKQGHFEIILKISVNSKNAKYVYGNNTKSIETKATKYDPHSKTMTPDLYANIN 180
 DB 1433 YGVLSPSKQGHFEIILKISVNSKNAKYVYGNNTKSIETKATKYDPHSKTMTPDLYANIN 1492
 QY 181 YVDGLAFAGDMRLFVNDQDKAEIKIRMEPIKETKSEYVYSSYGVNIELGEGDLSK 240
 DB 1493 YVDGLAFAGDMRLFVNDQDKAEIKIRMEPIKETKSEYVYSSYGVNIELGEGDLSK 1552
 QY 241 KPNULTKMSGKTYSEKQYILKONIIIRKGYALKVTTYNKGTDMLEGNVYSKEDI 300
 DB 1553 KPNULTKMSGKTYSEKQYILKONIIIRKGYALKVTTYNKGTDMLEGNVYSKEDI 1612
 QY 301 AKIQKAMPNLRALSETTIYADSRNVEDGRSTQSVLSALDPFNIIRYQVTFPQNDKGEA 360
 DB 1613 AKIQKAMPNLRALSETTIYADSRNVEDGRSTQSVLSALDPFNIIRYQVTFPQNDKGEA 1672
 QY 361 IDKQGNLVTDSSKLVLFQKDKKEYTGBDKFVNAIKEDGSMPLIDTRPVNLSDMKYFNP 420

DB 1673 IDKQGNLVTDSSKLVLFQKDKKEYTGBDKFVNAIKEDGSMPLIDTRPVNLSDMKYFNP 1732
 QY 421 SKSNKIYVNNPEFYLRGKISDKGFWMLRVNESVVDNLYIGDLHIIDTR-DEFNKLVN 479
 DB 1733 SKSNKIYVNNPEFYLRGKISDKGFWMLRVNESVVDNLYIGDLHIIDTRDPFNKLVN 1792
 QY 480 KQDIDMDGMKDYKANFPDKYTDMDGNVYLTQTSYDLNAAKAVGVYQFLYNVKNDEVNI 539
 DB 1793 KQDIDMDGMKDYKANFPDKYTDMDGNVYLTQTSYDLNAAKAVGVYQFLYNVKNDEVNI 1852
 QY 540 DPKGNTSIYADGKSVFENINDKRNNGFDEIIOBHITVINGKRYTSFNDIKOIIDKTLNI 599
 DB 1853 DPKGNTSIYADGKSVFENINDKRNNGFDEIIOBHITVINGKRYTSFNDIKOIIDKTLNI 1912
 QY 600 KIVVDFPANTVKEIILNKDGEVSELKPHVVTVTIQNGKMSSTVSEEDFILPVYNG 659
 DB 1913 KIVVDFPANTVKEIILNKDGEVSELKPHVVTVTIQNGKMSSTVSEEDFILPVYNG 1972
 QY 660 ELEKGYQPDGWEISGFEKGDAGVYVNLSDKDTFIKPVFKIEEKEENKPTFDVSKKD 719
 DB 1973 ELEKGYQPDGWEISGFEKGDAGVYVNLSDKDTFIKPVFKIEEKEENKPTFDVSKKD 2032
 QY 720 NPQVNSQLNESHKREDLQREHSHQSKSDTKVATVLDKNNISKSTNNPNK 773
 DB 2033 NPQVNSQLNESHKREDLQREHSHQSKSDTKVATVLDKNNISKSTNNPNK 2086
 RESULT 6
 ADK48759
 ID ADK48759 standard; protein; 2138 AA.
 XX
 AC ADK48759;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE Streptococcus pneumoniae protein, Seq ID No 5274.
 XX
 KM Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN US6699703-B1.
 XX
 PD 02-MAR-2004.
 XX
 PF 26-MAY-2000; 2000US-00583110.
 XX
 PR 02-JUL-1997; 97US-0051553P.
 PR 12-MAY-1998; 98US-0085131P.
 PR 30-JUN-1998; 98US-00107433.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Doucette-Stamm J, Bush D, Zeng Q, Opreman T, Housewart CE;
 XX
 DR WPI: 2004-212399/20.
 DR N-PSDB; ADK46098.
 XX
 PT New nucleic acid molecules and polypeptides useful for diagnosing,
 PT preventing and treating pathological conditions resulting from bacterial
 PT infection, e.g. Streptococcus pneumoniae infection, and in drug
 PT screening.
 XX
 PS Disclosure, SEQ ID NO 5274; 301pp; English.
 XX
 SS The invention relates to isolated Streptococcus pneumoniae nucleic acids
 CC and polypeptides. The nucleic acids and proteins are useful for
 CC diagnosing, preventing and treating pathological conditions resulting
 CC from bacterial infection, such as S. pneumoniae infection. These may also
 CC be used for drug screening procedures. The present sequence represents a
 CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
 CC data for this patent did not appear in the printed specification but was

CC obtained in electronic format directly from USPRO at
CC segdata.uspto.gov/sequence.html.

XX Sequence 2138 AA;

Query Match 94.1%; Score 3789.5; DB 8; Length 2138;
Best Local Similarity 94.8%; Pred. No. 1.5e-210;
Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

```

QY 1 KLGEIAESKPKNLGNKESGLKKDVTGVHHHQQEENESIKESKSPFIDENISTIDFENK 60
DB 1333 KLGEIPESKPKNLGNKESGLKKDVTGVHHHQQEENESIKESKSPFIDENISTIDFENK 1392
QY 61 DLKGLIKKKFREVDFSTSTGKMEBYDYKTDKGNIIAYDGTDLLEYTEKLDSEIKSI 120
DB 1393 DLKGLIKKKFREVDFSTSTGKMEBYDYKTDKGNIIAYDGTDLLEYTEKLDSEIKSI 1451
QY 121 YGVLSPSKDGHEILIGKISNVSKNAVYVYNNYKSIETATYDHPFSKTMETDLVANI 180
DB 1452 YGVLSPSKDGHEILIGKISNVSKNAVYVYNNYKSIETATYDHPFSKTMETDLVANI 1511
QY 181 IYDGLAFAGDMRLFVNDNDOKTAETIKIRMEPIKETSEYPIVSYSGNVIIEGEGDLSKN 240
DB 1512 IYDGLAFAGDMRLFVNDNDOKTAETIKIRMEPIKETSEYPIVSYSGNVIIEGEGDLSKN 1571
QY 241 KPDNLTMSGKIYSDSEKQYLLKDNIIIRKGYALKVTTPNGKTDMLENGVYSKEDI 300
DB 1572 KPDNLTMSGKIYSDSEKQYLLKDNIIIRKGYALKVTTPNGKTDMLENGVYSKEDI 1631
QY 301 AKIQKAPMLRALSETTIYADSRNVEDGSTQSVLMSALDGFNIIRYOVFTFRKMDKGA 360
DB 1632 AKIQKAPMLRALSETTIYADSRNVEDGSTQSVLMSALDGFNIIRYOVFTFRKMDKGA 1691
QY 361 IDKDGMLVTDSSKLVLFKGDDEKYTEGDEKFNVEAIKEDGSMFLIDTKPVNLSMDKNYFNP 420
DB 1692 IDKDGMLVTDSSKLVLFKGDDEKYTEGDEKFNVEAIKEDGSMFLIDTKPVNLSMDKNYFNP 1751
QY 421 SSKNKIYVNNPFIYLRGKISDYGKGFNWEIRVNESVVDNYLIIGDLHIDTRDFNKLANK 480
DB 1752 SSKNKIYVNNPFIYLRGKISDYGKGFNWEIRVNESVVDNYLIIGDLHIDTRDFNKLANK 1811
QY 481 DDIDIMGMKDYKANGFPDKVTMDGNVYLOTGYSDDLAKAVHNYQPLVYNDKPEVND 540
DB 1812 DDIDIMGMKDYKANGFPDKVTMDGNVYLOTGYSDDLAKAVHNYQPLVYNDKPEVND 1871
QY 541 PKGNTSIEYADGKSVFENIDNRKNGFDEIOEHIIYINGKETSFNIDIKIIDKTLNIX 600
DB 1872 PKGNTSIEYADGKSVFENIDNRKNGFDEIOEHIIYINGKETSFNIDIKIIDKTLNIX 1931
QY 601 IYVKKPFRATTYKEFLANKDTGVSFLKPHRYVTIIONGKEMSTTVSEEDTILPYKGE 660
DB 1932 IYVKKPFRATTYKEFLANKDTGVSFLKPHRYVTIIONGKEMSTTVSEEDTILPYKGE 1991
QY 661 LEKGYOFDMEISGFEKGDAGVNLISDQTFIKPFFKIEEKKEENKPTDVSKKON 720
DB 1992 LEKGYOFDMEISGFEKGDAGVNLISDQTFIKPFFKIEEKKEENKPTDVSKKON 2051
QY 721 POWNHSHQSHRKEDLOREHSHQSDSTKYDTATVLDKNNISKSTNNPNK 773
DB 2052 POWNHSHQSHRKEDLOREHSHQSDSTKYDTATVLDKNNISKSTNNPNK 2104

```

RESULT 7
ADR94534
ID ADR94534 standard; protein; 637 AA.
XX ADR94534;
AC
XX
DT 16-DEC-2004 (first entry)
XX
DE Novel S. pneumoniae protein sequence, SEQ ID 3169.
XX
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;

KW bacterial infection.

XX Streptococcus pneumoniae.

XX US6800744-B1.

XX 05-OCT-2004.

XX 30-JUN-1998; 98US-00107433.

XX 02-JUL-1997; 97US-0051553P.

XX 12-MAY-1998; 98US-0085131P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX MPI: 2004-697205/68.

XX N-PSDB; ADR91931.

XX New isolated nucleic acid encoding a Streptococcus pneumoniae polypeptide, useful for diagnosing, preventing and/or treating pathological conditions resulting from the bacterial infection.

XX Disclosure; SEQ ID NO 3169; 151pp; English.

The invention relates to an isolated nucleic acid comprising a sequence encoding a Streptococcus pneumoniae ADR91366polypeptide, or its fragments, with any of 9 fully defined sequences (appearing as ADR94308, ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682, ADR96079) or any of the fully defined sequences appearing as ADR91705, ADR91886, ADR92157, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide sequences, or at least 40, 60 or 300 consecutive nucleotides, which is hybridizable under high stringency conditions to the nucleotide sequence. The nucleic acids and proteins are chosen from 5206 disclosed sequences. Also included are a recombinant expression vector comprising the isolated nucleic acid cited above operably linked to a transcription regulatory element, a cell comprising the recombinant expression vector and a probe comprising at least 20 consecutive nucleotides of the nucleotide sequences as cited above. The methods and compositions of the present invention are useful for the diagnosis, prevention and/or treatment of pathological conditions resulting from bacterial infection by Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and otitis media. The present sequence is one of the 2603 disclosed S. pneumoniae protein sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPRO at segdata.uspto.gov/sequence.html?docid=6800744B1.

XX Sequence 637 AA;

Query Match 75.7%; Score 3048; DB 8; Length 637;
Best Local Similarity 99.3%; Pred. No. 2.8e-168;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

QY 185 IAFAGDMRLFYKNDNDOKTAETIKIRMEPIKETSEYPIVSYSGNVIIEGEGDLSKNKPDN 244
DB 15 IAFAGDMRLFYKNDNDOKTAETIKIRMEPIKETSEYPIVSYSGNVIIEGEGDLSKNKPDN 74
QY 245 LTKMESGKIYSDSEKQYLLKDNIIIRKGYALKVTTPNGKTDMLENGVYSKEDIATIQ 304
DB 75 LTKMESGKIYSDSEKQYLLKDNIIIRKGYALKVTTPNGKTDMLENGVYSKEDIATIQ 134
QY 305 KANPMLRALSETTIYADSRNVEDGSTQSVLMSALDGFNIIRYOVFTFRKMDKGAIDKD 364
DB 135 KANPMLRALSETTIYADSRNVEDGSTQSVLMSALDGFNIIRYOVFTFRKMDKGAIDKD 194
QY 365 GNLVTDSSKLVLFKGDDEKYTEGDEKFNVEAIKEDGSMFLIDTKPVNLSMDKNYFNPSSKN 424
DB 195 GNLVTDSSKLVLFKGDDEKYTEGDEKFNVEAIKEDGSMFLIDTKPVNLSMDKNYFNPSSKN 254
QY 425 KIYVNNPFIYLRGKISDYGKGFNWEIRVNESVVDNYLIIGDLHIDTRDFNKLANKDGD 484

```

```

|||||
255 KTVNPEFYLKRLKSDKGFNWELEFVNSVDNYLIGDLHIDNRDPNKLNVQDGI 314
485 MDGMGMDYKANGPPDKVTDMDGNVYQOTGSDLNAAKAVGHVQFLYDNKPEVNIIDPKGN 544
315 MDGMGMDYKANGPPDKVTDMDGNVYQOTGSDLNAAKAVGHVQFLYDNKPEVNIIDPKGN 374
545 TSIEYADGKSVFNIIDKNNNGFDEIIOEHYIYNGKEYTSFNDIKQIIDKTLNIIYVK 604
375 TSIEYADGKSVFNIIDKNNNGFDEIIOEHYIYNGKEYTSFNDIKQIIDKTLNIIYVK 434
605 DPAKNTTVEKFLNKDTGVESELEKPHRVYTTIONGKEMSTIYSEEDFILPYKGELENG 664
435 DPAKNTTVEKFLNKDTGVESELEKPHRVYTTIONGKEMSTIYSEEDFILPYKGELENG 494
665 YQPDGWEISGPEGKAGVYINLSKDTFLKPVFKIIEKKKEENKPTPVSKKKNPOVN 724
495 YQPDGWEISGPEGKAGVYINLSKDTFLKPVFKIIEKKKEENKPTPVSKKKNPOVN 554
725 HSQLESNRKEDLOREESQKSDSTKDVATVATVLDKNNISKSTNNPNK 773
555 HSQLESNRKEDLOREESQKSDSTKDVATVATVLDKNNISKSTNNPNK 603

RESULT 8
ADR96136 standard; protein; 1529 AA.
XX ADR96136;
AC ADR96136;
DT 16-DEC-2004 (first entry)
XX
DE Novel S. pneumoniae protein sequence, SEQ ID 4771.
XX
KM Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
XX bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN US6800744-B1.
XX
PD 05-OCT-2004.
XX
PF 30-JUN-1998; 98US-00107433.
XX
PR 02-JUL-1997; 97US-0051553P.
XX 12-MAY-1998; 98US-0085131P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI, 2004-697205/68.
XX N-PSDB; ADR93533.
XX
PT New isolated nucleic acid encoding a Streptococcus pneumoniae
XX polypeptide, useful for diagnosing, preventing and/or treating
XX pathological conditions resulting from the bacterial infection.
XX
PS Disclosure, SEQ ID NO 4771; 151pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising a sequence
XX encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
XX fragments, with any of 9 fully defined sequences (appearing as ADR94308,
XX ADR94488, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
XX ADR96079) or any of the fully defined sequences appearing as ADR91705,
XX ADR91886, ADR92197, ADR92234, ADR93039, ADR92366, ADR92650 or
XX ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
XX sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
XX hybridisable under high stringency conditions to the nucleotide sequence.
XX The nucleic acids and proteins are chosen from 5206 disclosed sequences.
XX Also included are a recombinant expression vector comprising the isolated
XX nucleic acid cited above operably linked to a transcription regulatory

```

```

CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=6800744B1.
XX
SQ Sequence 1529 AA;
XX
Query Match 18.6%; Score 750.5; DB 8; Length 1529;
Best Local Similarity 79.7%; Pred No. 1,3e-34;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;
XX
QY 1 KLGEIAESKFKNLGNGKEGSLKDDTGVVHHHQBNEESIKESPTIDNISTIRDFENK 60
DB 1336 KLGEIPESKFKNLKXVQKDSLNKETAIVENNLVNOGSIKGLFNIHKTISTIRDFENK 1395
QY 61 DLKLLIKKKFREVVDFTSETGKMEBYDYKDKGNIIAYDGTDLFETEKLDKIKSKI 120
DB 1396 DLKLLIKKKYKQEDDFVN-GGRTVERDYKDKGNIIAYDGTDLFETEKLDKIKSKI 1454
QY 121 YGVLSPSKDGHEIIGKISNVGSKNAKVYGGNNYKSIKATKYDFSKTMTEDLVANIND 180
DB 1455 YGVLSPSKDGHEIIGKISNVGSKNAKVYGGNNYKSIKATKYDFSKTMTEDLVANIND 1514
QY 181 IVDGLAF 187
DB 1515 IVDGISF 1521

RESULT 9
AAMS5096
ID AAMS5096 standard; protein; 117 AA.
XX
AC AAMS5096;
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0043 protein.
XX
KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US019422.
XX
PR 31-OCT-1996; 96US-0023960P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX
DR WPI, 1998-272224/24.
XX N-PSDB; AAV27357.
XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX - or their epitope-containing fragments, useful in protective or
XX therapeutic vaccines, and for diagnosis.
XX
PS Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein

```

CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX
SQ Sequence 117 AA;

Query Match 15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.2e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKVPFKIEKKEENKPTFDVSK 716
DB 1 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKVPFKIEKKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHRKEDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 773
DB 61 KKDNPQVNHSQLNESHRKEDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 117

RESULT 10

ABP54590
ID ABP54590 standard; protein; 117 AA.

AC ABP54590;

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

KW antibacterial; Streptococcal infection; detection.

OS Streptococcus pneumoniae.

XX US2002061545-A1.

PD 23-MAY-2002.

PF 22-JAN-2001; 2001US-00765272.

PR 30-OCT-1997; 97US-00961083.

XX (CHOI/) CHOI G H.

PA (KUNS/) KUNSCH C A.

PA (BARA/) BARASH S C.

PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.

PA (FANN/) FANNON M R.

PA (ROSE/) ROSEN C A.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

PI Rosen CA;

XX WPI; 2002-479261/51.

DR N-PSDB; ABQ84825.

XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus

PT and for preventing or attenuating disease caused by Streptococcus

XX infection.

XX Claim 11; Page 29; 70pp; English.

CC ABQ84792 to ABQ84904 represents nucleic acids which encode the

CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.

CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcal infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention
XX
SQ Sequence 117 AA;

Query Match 15.3%; Score 615; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.2e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKVPFKIEKKEENKPTFDVSK 716
DB 1 YKGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKVPFKIEKKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHRKEDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 773
DB 61 KKDNPQVNHSQLNESHRKEDLQREHRSQKSDSTKVATVLDKNNISSKSTNNPNK 117

RESULT 11

ADC45149
ID ADC45149 standard; protein; 117 AA.

AC ADC45149;

DT 18-DEC-2003 (first entry)

DE S. pneumoniae antigenic protein SP043.

XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.

KW Streptococcus pneumoniae.

OS US6573082-B1.

XX 03-JUN-2003.

PF 28-MAR-2000; 2000US-00536784.

PR 31-OCT-1996; 96US-0029960P.

XX 30-OCT-1997; 97US-00961083.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

PI Rosen CA;

XX WPI; 2003-764574/72.

DR N-PSDB; ADC45148.

XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides

PT useful for producing vaccines for prevention or attenuation of infection

XX by Streptococcus pneumoniae.

XX Example 1; SEQ ID NO 68; 58pp; English.

XX The invention relates to an isolated polynucleotide consisting of a

CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding

CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae

CC antigens. Also included are making a recombinant vector by inserting the

CC nucleic acid into a vector, an isolated polynucleotide consisting of at

CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a

CC recombinant host cell comprising the SP028 polynucleotide. The nucleic

CC acids are useful as DNA vaccine against Streptococcus pneumoniae

CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae

CC antigen nucleic acids are useful as probes for use in diagnostic methods

CC for detecting S. pneumoniae gene expression. The present sequence

XX represents an S. pneumoniae antigenic protein.

Seq Sequence 117 AA;
Query Match 15.3%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.2e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 657 YKGELEKGYQPDGWEISGFEKKDAGVYVNLKSDTFIKVFKKIEKKEENKPTDVS 716
DB 1 YKGELEKGYQPDGWEISGFEKKDAGVYVNLKSDTFIKVFKKIEKKEENKPTDVS 60
QY 717 KQDNQVNHNSQVNSRKKEDLQREHSQKSDSTKDTATVLDKNISSTNNPNK 773
DB 61 KQDNQVNHNSQVNSRKKEDLQREHSQKSDSTKDTATVLDKNISSTNNPNK 117
RESULT 12
ABU24813
ID ABU24813 standard; protein; 1196 AA.
AC ABU24813;
XX
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by Prokaryotic essential gene #10340.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Clostridium botulinum.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-0007851.
XX
XX 06-MAR-2002; 2002US-0362599P.
XX
XX (ELIT-1) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,
XX
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA28683.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX
XX PT for homologous nucleic acids required for cellular proliferation to
XX
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 52737; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX
XX CC the 6213 antisense sequences given in the specification where expression
XX
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX
XX CC polypeptide or its fragment whose expression is inhibited by the
XX
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX
XX CC proliferation; (7) identifying a gene in an operon required for
XX
XX CC proliferation; (8) identifying a gene that influences the activity of
XX
XX CC the gene product or that has an activity against a biological pathway
XX
XX CC required for proliferation, or that inhibits cellular proliferation; (9)
XX
XX CC identifying a gene required for cellular proliferation or the biological
XX
XX CC pathway in which a proliferation-regulated gene or its gene product lies
XX
XX CC or a gene on which the test compound that inhibits proliferation of an
XX
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX
XX CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIP0 at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Seq Sequence 1196 AA;
Query Match 5.6%; Score 227; DB 6; Length 1196;
Best Local Similarity 20.2%; Pred. No. 0.00021;
Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;
QY 1 KLGIAESKFKNLGNGKSG-----SLKDDTGVHGHQENES 39
DB 123 RLKDIQE-LFMDTGIGKSGYSIIIGOKIEAULSGPERRSLLEBAAGIVKRMKEAD 181
QY 40 KEKSFITDRNISTIRDPEN--KDLKLIKKKFREVD--TSETGRME----- 85
DB 182 KKLGN--TEQNLIRKIDILNTEBERMEPEKSEKAKFLNSELKREVVMIYSIDK 239
QY 86 -EYKRYDDKNIIVYDGD--LEYETKLDIKSIYGVLSPSDGHFEILGKISNYSK 143
DB 240 IEKDLK-NISSNMLISGENIDMLKNEKSGYKEIISKF-----NEKLELDK--NNSR 288
QY 144 NAKVYGNVYKSIIEIKATFYDFHSHKTMFEDLVANINDIVGLAFAGDMRLPYKDDQCKA 203
DB 289 NKEBYTNKDKNKDIE-----NENALKEKIKLND-----NIKVENLKTN 331
QY 204 EIKIR-----MEKIKETSEYYPV----- 223
DB 332 EERLQKYLKKEKALEKINKLKEESYLRBEIKGKEDNVNNPNKELKEEKILKISRE 391
QY 224 -----SSYGNVIELGEGDL--SNKPDNLTTMESGKIYS-----DSEKQOYLAK 265
DB 392 IEILNNTSNLRNEISVMEIEIVNLKNDIKNSCDSYSSININIKTKEDIEKEIKNIK 451
QY 266 DNIIIL-----KGYALKTYTNPQKTDMLBENGVSXED-----IAKIOK----AN 307
DB 452 ENILLLENNKENSNNISLKTSLNNKEX-KLKEKNAAYSRLBANVHMLSNLEKHYEGYN 510
QY 308 PNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFPMNDKGEAIDKGNL 367
DB 511 RSVKTLMEHVSYGKVDNIKGS-----CEVLGDIIVKXKELEFAMEIALGAI---SNV 560
QY 368 VTDSKLVLFGRKDKYTGEDKPNVAIKEDSGMLFIDTKPVNLSMDKNYFNP--SKSN 424
DB 561 ITEDE-----NKAKILINYLKCK-----SLGRATFPLPTTIQOR 594
QY 425 KIYVNPPEFYLRGKISDKGFWMLRVNSESVDNVLIVGDLHIDMTDRDNI--KLNVKXG 482
DB 595 KAKINN-----VTRDGF--GLIASDLIDY-----DVFSNIDIVLGRTLAKO- 637
QY 483 DLMWGMDKYGKANGFPDKYTDMDGNVYLQGTYSDNAKAVGHVQFLYDNVKPEVNI 542
DB 638 --MDSALKIATKLANISFKIVLEGV-----ING 665
QY 543 GNT---SIEYADGKSVENINDKRNNGPDGEIOEHYIYNGKEYSFNDIKQIIDKTLNI 599
DB 666 GSLTGSGISGRAGSSI--ISRR-----EIEE-----TKLELETKN- 700
QY 600 KIVVDPANVTYKPEILINKDGEVSELKPHVVTYTIQNGKEMSSYVSEDPFILPVYVG 659
DB 701 -----TIEEWGN-----ILEKNKIKY--LDEENLNI---KD 728
QY 660 ELEKGYQPDGWEISGFEKKDA-----GVINLSKD--TFIKVPFKKIEKKEENK 709

PR 27-AUG-2002; 2002US-040676P.
 PR 28-AUG-2002; 2002US-0406757P.
 XX
 PA (CHTR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Tetelijn H, Maignani V;
 XX
 DR WPI; 2004-248071/23.
 XX
 PT Immunogenic composition useful as a vaccine for treating or preventing
 PT streptococcal infections, comprises group B streptococcus polypeptides.
 XX
 PS Claim 10; SEQ ID NO 4422; 1194pp; English.
 XX
 CC The invention relates to a novel immunogenic composition comprising a
 CC combination of 2-5 group B streptococcus (GBS) polypeptides. Each
 CC polypeptide is encoded by a GBS polynucleotide sequence which is
 CC homologous to a polynucleotide sequence of group A streptococcus (GAS),
 CC streptococcus pneumoniae and/or at least one other GBS serotype. The
 CC composition of the invention demonstrates antibacterial activity whilst
 CC the polypeptides and polynucleotides may be useful in assays to diagnose
 CC and identify streptococcal infections or for identifying, screening and
 CC developing vaccines and other treatments for streptococcal infections.
 CC The current sequence is that of a streptococcus agalactiae ORF 5AG
 CC protein of the invention.
 CC
 SQ Sequence 1166 AA;

Query Match 5.5%; Score 220; DB 8; Length 1166;
 Best Local Similarity 20.2%; Pred. No. 0.00051;
 Matches 198; Conservative 139; Mismatches 306; Indels 336; Gaps 56;

QY 4 ELAESKFKNLGNGKSGSLKDTTGVHHQENBESIKESFTIDRINISTIDFENKDK 63
 DB 60 ELLELSKSLDTSNMGK-----DLSEYPSKPEYTNNEES-NVVTNASTA-----IA 105
 QY 64 KLIIKKFREVDFTS-----ETGKMEBYDYKYDDKGNIIA-YDDGTLEYETEKLDDEI 116
 DB 106 QKVPAAVEYEVKPKSSSLAVLDTSKITKLQATQKRGKVVALLIDGPIINIDIPDL-- 163
 QY 117 KSKIYGLVSPSKDGH-----FEIIGKISNVS---KNAKYVYGNKYK-----TEIKA 160
 DB 164 -----SPKDDKISFKKKAFFEBELKAKGNITYGKRVNDKIVFANVANNITETVADIAA 215
 QY 161 TKYDFHSKMTFDLVANINDIVDGLAFAGDMFLFYVDNOKKA-----EIK 206
 DB 216 AMKDGIG-----SEAKNISHGTHVAG--IFV-GNSKRPAINGLLLEGAAPNAQVLL 263
 QY 207 IMPEKIKETKGEYPPVSSYGVNIELEGDLISKNKPDNLTMSKSGK-----IYDSSEKQOY 262
 DB 264 MFIPIKIDSKFGEAYAKAITDAVNLGAKTIN-----MSIGTASLSLALANDKVKL 314
 QY 263 LKNDIILKGYALKVTTYNPGKTDMLEGNGVYSKEDIKIQANP-----NLRALSETT 317
 DB 315 ALK--LASEKGVAVVVAAGNEGAFGM-----DYSKPLSTNDYGTVNSPAISEDT 362
 QY 318 IYADSRNVDGSGTQSVLMSALDGFNIIYQVTFPMNDKGAIDGDGVLVDSGLVLF 377
 DB 363 LSVAS--YSLKTISEVETTIEG-KLVKLPVTSFPPKGRAYD-----VVY 407
 QY 378 GK--DDKEYTGED-KFNVAIKEDGSMFLID--TKVNLSDMDKNVFPNSKNKIYVRNPE 432
 DB 408 AAYGAKKDEFGKDFPKKALIERGGLDPMTKITHTATNAGV-----VGIYVENQO 457
 QY 433 FYLRGKISDKGFGFMELRVNESVVDNYLI-YGDL-----HIDNTRD--FNIKL 477
 DB 458 -----EKRG-----NFLIPYRELPGVVISKYDGEIKNTKTSQLTFFNQF 496
 QY 478 NYKD--GDIM-----DWGM-----KDYKANGFPKXVTDMDGNVYLVQGYGDLANKKAG 523
 DB 497 EYVDSGGNRMDEQSSWGVYTAGCAIKPVDYASGEFIIYSSTYNNQYQYTMGTSMASPHVAG 556

QY 524 V-----HYOFLYDNVKEPVNIDPK-----GNTSIEYADGKSVFNFINDKRNNG 566
 DB 557 LMTMLQSHLAEKYKG-----MMLDSKGLLELSKNILMSATLYSEEDKAFYSPRQOGAGV 612
 QY 567 FDGE--IOEHIIYNGKSYTSPNDIKQIIDTINKIYVVKDPARTTYKEFLINKDTGEV 624
 DB 613 VDAEKAIQAQY-YVVGNDGKAKIMKRVGDK-FDITVTHIMLVESGKELYYOANVATROV 670
 QY 625 SE-----LKPW-----RVVTYIQ-----NG----- 639
 DB 671 NKGKALRFOALLDNTNMOKVILRDKETOVRFITDASQSKLKEQMANGYFLEGVREKE 730
 QY 640 -KEMSGTIVS-----BEDFI-----LPVYKGELEKXY-----QFDGMEISG 674
 DB 731 AKDSNGELMSIPFVGNDGDFASLQALETPIYK-TLSKSPFYKPNDDTHKQLEYNESAP 789
 QY 675 FEGRK-----DAGIV-----INLSKDT--FIKPYK-KIEK-----KEBE 707
 DB 730 FESNNYTLALTQSASGWYVDYVYKNGEELAPESPRIILGTFENKVEDKTIHLERDAA 849
 QY 708 NKPTFDVSKKDNPPQVNHNSQLNESHKEDLOREESHOKSD--STKDYATVLYDKN-NI- 762
 DB 850 NNPTFAISPNDG-----NDEITPQATFLRNKDISAOVLDQNGNVI 892
 QY 763 -----SSKSTTNPNK 773
 DB 893 WQSKVLPSTYRKNFHNNPKQ 911

RESULT 15

AAM54145
 ID AAM54145 standard; protein; 1639 AA.

AC AAM54145;

DT 25-MAR-2003 (revised)
 DT 23-SEP-1998 (first entry)

DE P. falciparum synthetic gp190 protein.

KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 KM monoclonal antibody; passive immunisation; parasite.

OS Plasmodium falciparum.
 OS Synthetic.

PN W09814583-A2.

PD 09-APR-1998.

PF 02-OCT-1997; 97WO-BP005441.

PR 02-OCT-1996; 96DE-01040817.

PA (BUA/) BUIYARD H.

PI Buiyard H, Tolle R, Pan W;

DR WPI; 1998-240089/21.

DR N-PSDB; AAV21451, AAV35363.

PT Recombinant production of complete gp190/MSP-1 Plasmodium surface protein
 PT - useful in anti-malaria vaccines, also stabilising genes by reducing
 PT their AT content.

PS Example 1; Fig 3c; 48pp; German.

CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface) protein. The gene encoding this protein has been
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 CC protein is useful in vaccines against malaria or for producing monoclonal
 CC antibodies (for passive immunisation). The complete gp190 protein can now
 CC be produced outside the parasite and has, at least over extended regions,

Db 1632 AKIQANPVLRLSETTIYADSRNVEDGRSTQAVLMSALDGNIRIYQVFTFRKANDKGEA 1691
Qy 361 IDKDNLTVDSSKVLVFGKDKREYTGEDKFNVEAIKEDGSMFLIDTKPVNLSDKNRYFNP 420
Db 1692 IDKDNLTVDSSKVLVFGKDKREYTGEDKFNVEAIKEDGSMFLIDTKPVNLSDKNRYFNP 1751
Qy 421 SSKNLTIVRNPEFYLRGKISDGKGFNWEILRVNESVVDNLTLYGDLHIDTRFNLKLVNK 480
Db 1752 SSKNLTIVRNPEFYLRGKISDGKGFNWEILRVNESVVDNLTLYGDLHIDTRFNLKLVNK 1811
Qy 481 DGDIMDMGKDYKANGFPDKTVDMDGNVYLQTYGSDLNKAGVHYQFLYDNVKEPVNID 540
Db 1812 DGDIMDMGKDYKANGFPDKTVDMDGNVYLQTYGSDLNKAGVHYQFLYDNVKEPVNID 1871
Qy 541 PKGNTSIEYADGKSVFVNINDKRNNGPDGEIOEHIIYNGKEYTSFNDIKQIIDTKLVNK 600
Db 1872 PKGNTSIEYADGKSVFVNINDKRNNGPDGEIOEHIIYNGKEYTSFNDIKQIIDTKLVNK 1931
Qy 601 IVVKDPARTTYKEFLINDTGEVSELKPHRTVTYTIQNGKEMSTTVSEDFILPVYKGE 660
Db 1932 IVVKDPARTTYKEFLINDTGEVSELKPHRTVTYTIQNGKEMSTTVSEDFILPVYKGE 1991
Qy 661 LEKGYPDGMELISGEGKKDAGVYINLSKDTFIKPVFKKIEKKEENKPTFDVSKKDN 720
Db 1992 LEKGYPDGMELISGEGKKDAGVYINLSKDTFIKPVFKKIEKKEENKPTFDVSKKDN 2051
Qy 721 PQVNHSQLNESHKEDLQREHSHQSKSDSTKDYATATVLDKNNISKSXTNNPNK 773
Db 2052 PQVNHSQLNESHKEDLQREHSHQSKSDSTKDYATATVLDKNNISKSXTNNPNK 2104

RESULT 2
US-09-107-433-3169
; Sequence 3169, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID and AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: Protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169
Query Match 75.7% Score 3048; DA 4; Length 637;
Best Local Similarity 99.3% Pred. No. 5,6e-191;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 185 LAFAGDMFLFYKNDQKAAEIKIRPEKIKETKSEHYPVSSYGVNVELEGDLSNKKPDN 244
Db 15 LAFAGDMFLFYKNDQKAAEIKIRPEKIKETKSEHYPVSSYGVNVELEGDLSNKKPDN 74
Qy 245 LTKMESGKIYDSEKQOYLINDNIIIRGVALKTTTNPQKTDMLEGNGVYSKEDIATQ 304
Db 75 LTKMESGKIYDSEKQOYLINDNIIIRGVALKTTTNPQKTDMLEGNGVYSKEDIATQ 134
Qy 305 KANPULRALSETTIYADSRNVEDGRSTQAVLMSALDGNIRIYQVFTFRKANDKGEAIDKD 364
Db 135 KANPULRALSETTIYADSRNVEDGRSTQAVLMSALDGNIRIYQVFTFRKANDKGEAIDKD 194
Qy 365 GNLVYDSSKVLVFGKDKREYTGEDKFNVEAIKEDGSMFLIDTKPVNLSDKNRYFNP 424
Db 195 GNLVYDSSKVLVFGKDKREYTGEDKFNVEAIKEDGSMFLIDTKPVNLSDKNRYFNP 254
Qy 425 KIYVNPPEFYLRGKISDGKGFNWEILRVNESVVDNLTLYGDLHIDTRFNLKLVNKDGI 484
Db 255 KIYVNPPEFYLRGKISDGKGFNWEILRVNESVVDNLTLYGDLHIDTRFNLKLVNKDGI 314
Qy 485 MDGMKDYKANGFPDKTVDMDGNVYLQTYGSDLNKAGVHYQFLYDNVKEPVNIDPKGN 544
Db 315 MDGMKDYKANGFPDKTVDMDGNVYLQTYGSDLNKAGVHYQFLYDNVKEPVNIDPKGN 374
Qy 545 TSIEYADGKSVFVNINDKRNNGPDGEIOEHIIYNGKEYTSFNDIKQIIDTKLVNK 604
Db 375 TSIEYADGKSVFVNINDKRNNGPDGEIOEHIIYNGKEYTSFNDIKQIIDTKLVNK 434
Qy 605 DPAVUTTYKEFLINDTGEVSELKPHRTVTYTIQNGKEMSTTVSEDFILPVYKELKG 664
Db 435 DPAVUTTYKEFLINDTGEVSELKPHRTVTYTIQNGKEMSTTVSEDFILPVYKELKG 494
Qy 665 YQFDGMEISGEGKKDAGVYINLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNPOVN 724
Db 495 YQFDGMEISGEGKKDAGVYINLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNPOVN 554
Qy 725 HSQLNESHKEDLQREHSHQSKSDSTKDYATATVLDKNNISKSXTNNPNK 773
Db 555 HSQLNESHKEDLQREHSHQSKSDSTKDYATATVLDKNNISKSXTNNPNK 603

RESULT 3
US-09-107-433-4771
; Sequence 4771, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID and AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4771:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...1529
SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-09-107-433-4771
```

Query Match 18.6%; Score 750.5; DB 4; Length 1529;
Best Local Similarity 79.7%; Pred. No. 1.8e-40;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

```

1 KGEIAESKFNKNGSGSLKKDTGTGVHHHGHQENESIKESSTFTIDNINISIRDFENK 60
1336 KGEIPESKFNKNGSLKKDTGTGVHHHGHQENESIKESSTFTIDNINISIRDFENK 1395
61 DKKLIKRRREVDFSTSGKRMEEYDYKDDKGNIIAYDGTDLVEYTEKLDIEIKSKI 120
1396 DKKLIKRRREVDFSTSGKRMEEYDYKDDKGNIIAYDGTDLVEYTEKLDIEIKSKI 1454
121 YGVLSKSGKGFHLLKISNVSKNAVYVGNVYKSLFIRATKTDPSKMTPLVYANIND 180
1455 YGVLSKSGKGFHLLKISNVSKNAVYVGNVYKSLFIRATKTDPSKMTPLVYANIND 1514
181 IVDGLAF 187
1515 IVDGISF 1521
Db

RESULT 4
US-08-961-083-68
Sequence 68, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
```

```

OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Andreb
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-68

Query Match 15.3%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.9e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

657 YGGELEKGYQPDGWEISGFGKDAAGYVNLKDTFIKVPFKIEKKEENKPFVDVSK 716
1 YGGELEKGYQPDGWEISGFGKDAAGYVNLKDTFIKVPFKIEKKEENKPFVDVSK 60
717 KKDQVNVHSQUNESHREKDLQREHSQKSDSTKQVTAIVLADKNNISSKSTNNPK 773
61 KKDQVNVHSQUNESHREKDLQREHSQKSDSTKQVTAIVLADKNNISSKSTNNPK 117
Db
```

```

RESULT 5
US-09-536-784-68
Sequence 68, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
```

```

;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 117 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match          15.3%; Score 615; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.9e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWEISGFEKGDAGYVNLISKDTPIKVFKEIEKKEENKPTFDVSK 716
DB 1 YKGELEKGYQPDGWEISGFEKGDAGYVNLISKDTPIKVFKEIEKKEENKPTFDVSK 60
QY 717 KKDQNVNHSQNLNEHREKEDLQREHSSQKSDSTKVATVLDKNNISSKSTNNPNK 773
DB 61 KKDQNVNHSQNLNEHREKEDLQREHSSQKSDSTKVATVLDKNNISSKSTNNPNK 117

RESULT 6
US-09-200-650E-7
; Sequence 7, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
;   APPLICANT: Patti, Joseph M.
;   APPLICANT: Foster, Timothy J.
;   APPLICANT: Hook, Magnus A.O.
;   APPLICANT: Eidiham, Deirdre M.
;   APPLICANT: Perkins, Samuel L.
;   TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
;   FILE REFERENCE: P06283US2/BAS
;   CURRENT APPLICATION NUMBER: US/09/200,650E
;   CURRENT FILING DATE: 1998-11-25
;   PRIOR APPLICATION NUMBER: 60/066,815
;   PRIOR FILING DATE: 1997-11-26
;   PRIOR APPLICATION NUMBER: 60/098,427
;   PRIOR FILING DATE: 1998-08-31
;   NUMBER OF SEQ ID NOS: 23
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 7
;   LENGTH: 1166
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-200-650E-7

Query Match          5.0%; Score 203; DB 4; Length 1166;
Best Local Similarity 21.6%; Pred. No. 6.8e-05;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

QY 33 QENESIKE--KSFITIDNISTIRDFEKDLKLIKKEFR-----EYDDFTS 78
DB 226 KNNPEKLEIVANDSTHSTPAPATAPSAFKVNAKMRRAVAVAPAAVANNVNDLIX 285
QY 79 ETGKMEEDYVYDDKGNIIAVDGTLEETE-----KLDEIKSKIYGVIS 125
DB 286 VTQKQIK---VGDKDNVAHAHDGDIJEYDFEITDNKKKGDITMTIYDKGVIPSDLT 341
QY 126 PSKD-----GHFEILGISNVSKNAK---VYGNMYSKLEIKATKDHSK----- 168
DB 342 DKNDIDITDTPSGEVIKAGTFDKATKQITTYPTDVADEKEDIKSRLLIYSYIDKKTVPNE 401
QY 169 ---TWTF-----DLVANIN-DIVDGLAFAGD---RELFEVK-DNDQKAEIKIRM-PEKIX 214
DB 402 TSLNITFATAGKETGQNTVVDQDPNVH-GDSNIOISIFKLDEDEKQITIEQIYVNPPLKKS 460
QY 215 ETKSEY---PYVSYGVNIELEGED--LSKNKPPNLTQM-----ESGKIYSDSEKQO 261
DB 461 ATNTKYDIAGSQVDVYGN-IKLNGSGSTIIDQNTETIKVYKVNSDQQLPGSNRIYDPSQYED 519
QY 262 YLLK-DNIIIRKGYALKTTTYPGKTIDMEGNGVYSKEDIKIQANPMLALSTETIYA 320
```

```

DB 520 VTSQPDN---KKSPSNNAVATLDPG-----DINSAIT-----IKVSKYTPS 558
QY 321 DSR-NVEDGRSTQSVLMSALDGFNIRYQVTFKANDKG-----E 359
DB 559 DGEIDIAQGTSMRTT--DKGYVYVAGSYNSIVTNSNDVGGDGTVKPEEKLYKIGDYWE 616
QY 360 AIDKDNLVTSKSLVLFQKDKDEYTGDKFNVEAIKEDGSMFLIDYKPVNLSMDKNY-F 418
DB 617 DVDKDG-----VGDTDSKEKPMANVL-VTLTYPDGT-----TKSVRTDANGHYEF 660
QY 419 NPSKSNKIYV---RNPEFYLNKGIS-----DKGFFWELRNVESVVDNVLIIYDGLHD 468
DB 661 GGLKDGETTYVKFETPTGYLPTKVNQTTDGEKDSNGSSVYKINK-----DDMSLD 712
QY 469 -----NTRDENIKILNVDG--DIMDMQKDYKANGPDKYVDMGNVYLQTVGSDL 517
DB 713 TGFYKPKRYNLDGYWEDTNDGIDQANPEIKQVKT-----LKDSIGKV-IGTTTIDA 766
QY 518 NAKAVGVHYQFLYDNVKEPVNIDPKNTSIEY---ADGKSVFN--INDKRNPFGEIIO 572
DB 767 SGK-----YKF-----TDL-NGNYTVEFETPAGYTPYKNTTADDKDSNGD----- 807
QY 573 EGHIIYNGKEYTSFNDIKQIIDKTLNIIKIYVKDPAANT--TVKERI---LKKDGEVSEL 627
DB 808 -----TTGVADADNMTLD---RGFYKTPKYSGLDGYWYDSNKGKQDSTE 851
QY 628 KPHR-VTVTIQNGK-EMSSITVSEEDFILPYKGELEKGYQPDGWEISGFEKGDAGYVI 685
DB 852 KGIDQVTVTLQNEKEVIGTKIDEN-----GK-----YFSD----- 883
QY 666 NLSKDTFIKVPFKIIEKKEENKPTFPVSKKQNPQV---NHSQ--LNEHREKEDLQ 738
DB 884 NLDSEK-KY-VIFEPKAGLQVTVTTED-DKDADGGEVDVITDHDFTLNGYFEEDTS 941
QY 739 REHSQKSDSTKD 751
DB 942 DSDSDSDSDSDSD 954

RESULT 7
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;   APPLICANT: DRUHLER, PIERRE
;   APPLICANT: DAUBERSIES, PIERRE
;   TITLE OF INVENTION: MALARIAL PRE-BRYTHOCYTIC STAGE POLYPEPTIDE MOLECULES
;   FILE REFERENCE: 0660-0125-0 PCT
;   CURRENT APPLICATION NUMBER: US/08/973,462B
;   CURRENT FILING DATE: 1998-02-06
;   EARLIER APPLICATION NUMBER: PCT/FR96/00894
;   EARLIER FILING DATE: 1996-06-12
;   EARLIER APPLICATION NUMBER: FR 95/07007
;   EARLIER FILING DATE: 1995-06-13
;   NUMBER OF SEQ ID NOS: 29
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 8
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match          5.0%; Score 200.5; DB 3; Length 1786;
Best Local Similarity 20.5%; Pred. No. 0.00018;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AES--KFRN-LGNGKEGSLKKDT-----TGVEHHQENE-----ESIK 40
DB 868 AESVTTFNIIIEIQENTITNDTIEKLELHENVLSAALNTQSEKKEKVIDVIEEVK 927
```


[illegible]

QY 316 TTYVADSNVBDGSGTOSVMSALDGNFIIRYQVTFPKNDKGALIDDGNLVYDSSKLV 375
Db 165 ATI-----LNGINIKNHPISIVFT-----GLFTDGAQV 193
QY 376 LFG-KDDKEYTGEDETFNEVAIKEDSGMLFIDTKPVNLMS--DKNYFNPSKRIYVRNP 432
Db 194 EMGFPEDISYSGGTIDNMGNALNEB-----TKANLPLINSSGAFALGNSNVTIKVYT 247
QY 433 F---YLRGKISDKGGFNNELRVNESVDVNYLIYGDLIIDNTRDENIKLVNKGDM--D 486
Db 248 FKDSYQGHAIQIAGSKN-----VLVNSRFLGA-----LPTMKDGGQIISKES 291
QY 487 WGMKDYKANGEPDKYTDMDG---NVLQTY---SDLNAK---AVGHHOELYONVKE 536
Db 292 IQIELETRKGPYALND-DGKSEBVTIIONSYFGKSDSGELVYAIQTHYQIOLTPONSN 350
QY 537 VNIDPKGNTSIEYADGKSVFNINDKRNNGFDEGIOEHY-----INGKEYTSFND 588
Db 351 IKILNNHDDNNMYAGVRFPTGFDVLIKGNRPDKKYGSGSVHYRESGALVYNAVYKNTKD 410
QY 589 I---KQII-----DYLNIKIYKOPAR-----NTYKEPLNKDTGVSSEIK 628
Db 411 LLDLAKQVVAIENTFNIADPKTKAIR-VAKDSAEYLGKVSITVYKVNINNSKTEQ-- 467
QY 629 PHRTVITIQNEKESSTIV--SEEDFILPVYKGEI---BKGYQDGEI-----SGFEGK 678
Db 468 FNIELLRSDLVVSENSIFGKGESIVIEDSKGKITVLNNGFYNLGKIYISFISNANGK 527
QY 679 K-----DAGVINLSKDTFIKPVFKKIEKKKEE-----NKPTFDVSK 716
Db 528 EPVIRDSGDNFNI-VTENGLYKIVTNLNSDKNEKKKNKEBKQYNSNNVYIDSNQKGEFNS 586
QY 717 KKNPQVNVHSQLNESHK-----EDIQREHSQKSTKDY 752
Db 587 SKDNKQNDKXIDNKQDNKTEBEVNYKIVGDGRETEHINKSKXIYDV 632

RESULT 10
US-09-583-110-3019
; Sequence 3019, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3019
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3019

Query Match 4.7%; Score 188.5; DB 4; Length 708;
Best Local Similarity 21.4%; Pred. No. 0.00031;
Matches 138; Conservative 92; Mismatches 233; Indels 183; Gaps 33;

QY 223 VSSYGVNVTLEGGDLSSKNKPNLTKMESGKIYSDSEKQY--LLKDNIIIR--KGVALK 277
Db 92 ISSNGTI-----RSNSQDLNRT-VES-TVTSITNENKSYKEDVYSDIHKKEFEDTALS 142
QY 278 VTTYN-----PGKTDMLSENGYISKEDIKIQ-----KANPNIRALISE 315
Db 143 VKDVGANVDGIHDDRQALQDAIDAAAGLGGGANYFPFGTYLVKEIVFLKSTHLELNK 202

QY 316 TTIVADSRNVEDGRSTQSVLMSALDGFNIIRYOVFTKRNKDGEALDKDGNLVTDSSKLV 375
 DB 203 ATIT-----LNCINIKNHSIYFMT-----GLFTDGAQV 231
 QY 376 LFG-KDKEVTGDEKFNVEAIKEDGSMFLDTKPVNLSN--DKVNPFSKNKIYVRNE 432
 DB 232 EWGPFEDISYSGGTIDMNGALNBEG-----TKAKVLPLINSAGALINSNVITIKVNT 285
 QY 433 F---YLKGISDKGFMELRVNESVVDVLLYGDHLINDTRDENIKLVKOGDLM--D 486
 DB 286 FDSYQGHAIQIAGSKN-----VYDNRPLGQA-----LPKTKGQQLISKES 329
 QY 487 WGMKQVKANPPDKVTDMDG-----NYLQTYG---SDLNK---AVGVHYQFLYDVKPE 536
 DB 330 IQIEPLTRKGFYALAND--DKKSENVITQNSYFGSKDSGELVTAIGTQITSTQNPEN 388
 QY 537 VNIDPKGNTSIEVADKSVVFNINDKRNNGFDEIQEHIY-----INGKEYTSFND 588
 DB 389 IKLNNHFDNNMYAGVRFGTQVLLKGNRFDKVGESVHYRESGALVNAVSYKNTKD 448
 QY 589 I-----KQIT-----DKTINIKIYKDFPKR-----NTYKFFILNKDTGEVSELK 628
 DB 449 LADLNKQVYAIENIFNIADPKTKAIR--VAKDSAEYLKVSQDITVTQVJNNNSKETEQ-- 505
 QY 629 PHRVTTIQNGKEMSSTIV--SEEDFILPVYKSEL-----EKGYQPDGWEI---SGFEGK 678
 DB 506 PNIELLRVDNLVNSNSIFGCKEIVIEDSKCKITVLANQFVNLISGKITSFKSNANK 565
 QY 679 K-----DAGVYINLSKDTFIKPVFKIEKKEE-----NKPTDVSK 716
 DB 566 EPVIRDSQNFNI--VTENGLYKLVTNLSDKNEKEKNEKQVNSNVVLDNQKNGEFPNS 624
 QY 717 KQDNQVNSQVNSHSHK-----EDLQREESQKSDSTKQV 752
 DB 625 SKDNQMDKIDNKQDNKTEEVNYKIVGGRTEHNNHNSKEIVDV 670

RESULT 11
 US-08-728-470-10
 ; Sequence 10, Application US/08728470
 ; Patent No. 5928651
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,470
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:
 ; NAME: Berckreaser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-633
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0810
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1529 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-728-470-10

Query Match 4.7%; Score 187.5; DB 2; Length 1529;
 Best Local Similarity 20.1%; Pred. No. 0.001;
 Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

QY 9 KFKNL---GNGKESLKKDTGVENHHOENESIKESFTIDRNISTIDPENKDLK-- 63
 DB 531 RNNVNSINGTK-----GLKEIANQNNTHKFDGELNIS--GIVTINQTKQDVKKW 580
 QY 64 KLIKKKFREVDPTSETGRMEEDYD--KYDDKGNIIAYDDGTDLEVEFEKLEIKSIYV 122
 DB 581 NASKOSYNNVSSLTINT--VQKFTFIKFDVSGS-----NSQDLSSRRSPAG 625
 QY 123 VLSPSKQGF--ILKGIS--NVSRAKAVY-----GNNYKSIET---KATKYPDHAKT 169
 DB 626 V-----HFNGIGKTFNIGANAKALFKLPNATDPKELPTTFNANITATGSDSS 678
 QY 170 MTFEDLYANIND-----IYDGLAFAGDMRLFVKNDOCKKAEIKIRMEPIKETSEVPYVS 224
 DB 679 VMFDIHANLTSAAIGIMDSITITGDLPSITSHRNSNAFEIKDQTLINAGSNN-- 735
 QY 225 SYGVNVLIEGDLISKNNKPNLTVMESGKIYSDSEKQVYLLKQNIILRKQVALKVTYTPNG 284
 DB 736 -----LKQTFDSFYNEYSKHALINSNHLRT----- 760
 QY 285 KTDMLEGNVYSKEDIQKANKPULKALSETTIYADSRNVEDGRSTQSVLMSALDGFNI 344
 DB 761 ---LGNVTLLEGNESSSITGNITNKNANVTLDATSNSTNGLKRRLLTLG----- 809
 QY 345 IRYQVFTFQMDKGEALIDKQNLVTDSSKLVLEK---DDKEYGDEKFNVEAIKEDGS 400
 DB 810 -----NISVEGNLSLTGANANIVGNLSIAEDSTPKGE----- 841
 QY 401 MLEFDTKPVNLSMDKNVFNPSKKNKIYVNPFEYLGKISDKGFMELRVNESVVDNYL 460
 DB 842 -----ASDNLNITGFTNNGTAN--INIKQGVVKLQGDINNKGGLN--ITTNASGTQKTI 892
 QY 461 IYGDHLINDTRDPNLIKLVK--DGDIMDG-----MKQYKAN-----G 496
 DB 893 INGNI--TNEKGDPLNIK--NIKADAEIQIGNISQKEGNLTSSDKVNTIQITIKAVEGG 950
 QY 497 PPDKVTMDMGVNYLOT-----GY-----SDL-----NAKAVGVHY 526
 DB 951 RSDSSEAEANALTIQKELKLAGDLNISFPNAKBITAKNGSDLTIGNASGADAKV-- 1008
 QY 527 QFLYDVKPEVNIIDPKGNTSIEVADKSVVFNINDKRNNGFD--GEIQEHIYINGKEYT 584
 DB 1009 ---TFDKVK-----DSKIST-----DGHNVTLNSEVTSNGSSNAGDNGSTGLTISAKDVT 1056
 QY 585 SPNDIKQIITDKTINIKIYVKDFARNTVVE--FILNKDTGEVSELKPHRTVYIONG----- 639
 DB 1057 VNNNVTS--HKTINISAA---AGNVYTKGFTINATGSVE-----VTAQNGTIFNG 1102
 QY 640 --KEMSTIVSEDFI 653
 DB 1103 NITSQNVYTAIENLV 1118

RESULT 12

US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: PC floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry M
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 4.7%; Score 187.5; DB 3; Length 1529;
Best Local Similarity 20.1%; Pred. No. 0.001; Indels 239; Gaps 37;
Matches 148; Conservative 94; Mismatches 255;

QY 9 KEKNTL---GNGKESLKKDTGVHHHNEESIKESFTIDRNISTIRDPENDLK-- 63
DB 531 RNNVNSINTGK-----GLKPIANONNFTHKFDGSLNIS-GVITINOTTKDVKW 580
QY 64 KLIKKKFREVDFTSETGRMBEYD-KYDDKGNIIAYDDGTDELEYETKLDIKSKIY 122
DB 581 NASKOSYMWVSLTNT---VQKFTPIKFDVDSG-----NSQDLSSRRSFAG 625
QY 123 VLSPEKDGHE-ILGKIS-NVSKNAKVY-----GNVYSIEI---KATKIDPHSKT 169
DB 626 V-----HFGIGGKTNNFIGNANALPKLPNATDKKELPTTFNANITATGSDS 678
QY 170 MTFEDLVANIND-----IYDGLAFAGDMRLFVNKDNDKKAIEKIRMEPIKETSEYFVS 224
DB 679 VWFIDHANLTSRAAGINMDSINITGSLDSITSHRNSNAFEIKDOLITNATGSNPS--- 735
QY 225 SYGVNIEIGEGDLSKNKPDNLTKMESGKIYDSEKQOYLKKDNIIILRKGYALKVTYNG 284
DB 736 -----LKQTKDSFYNEVSKAIINSNHLTI----- 760

QY 285 KTDMEGNGVYSKEDIKAKIQRANPILALSETTIYADSRNVEDGRSTQSLMADGFNI 344
DB 761 ---LGGNVTLGEGSSSSITGNITNKNKAVTLQADTSNSNTGKKRLLTLG----- 809
QY 345 IRYQVTFPMNDKGBAIDKGNLVTDSSKLVLFGR-----DDEKYTGDEKFNVEAIKEDGS 400
DB 810 -----NISVEGLSLTGANNANVGNLSIAEDSTFGE----- 841
QY 401 MLEIDTKPVNLSDMKVNFPSKINKIYVARNPEFYLRKISDKGFFNWEIYNESVVDNYL 460
DB 842 -----ASDNLNITGTFNNGTAN-INIKQGVAKQDGINNKGILN-ITTNASGTQKTI 892
QY 461 IYGDHIDNTDFNKLKWK-DGDIMDWG-----MKDYKAN-----G 496
DB 893 INGNI-TNEKGDNLNK-NIKADAEIQGNGISQKEGNLTSSDKVNTNQITKAGVEGG 950
QY 497 FPDKYTMDGNVYLOT-----GY-----SDL-----NAKAVGYH 526
DB 951 RSDSSEARNANLITITTKELKLAGDNLISGFNAEITAKNGSLTITGNASGGADAKV-- 1008
QY 527 QFLYNNVXPEVNIIDPKGNTSIEYADGKSVFNINDKRNNGFD--GEIQEHIYINGEY 584
DB 1009 --TFPKVK-----DSKIST-----DGHVTLINSVKTSGNSNAGNDNSTGLTISAKDVT 1056
QY 585 SFNDIKQIIDTKNLKIKYVKKOPARNTYKKE-FILKQTKGVSSELKPHRYTYTONG----- 639
DB 1057 VNNNVTG--HKTINISAA---AGNVTKEGTTINATGSAVE-----VTAQNGTITKG 1102
QY 640 --KEMSTIVSEEDPI 653
DB 1103 NITSGNVTVTATENLV 1118

RESULT 13
US-09-200-650E-3
Sequence 3, Application US/09200650E
Patent No. 6680195
GENERAL INFORMATION:
APPLICANT: Patti, Joseph M.
APPLICANT: Foster, Timothy J.
APPLICANT: Hook, Magnus A.O.
APPLICANT: Bidhni, Delidre N.
APPLICANT: Perkins, Samuel L.
TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
FILE REFERENCE: P06283052/BAS
CURRENT FILING DATE: 1998-11-25
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: 60/066,815
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: 60/098,427
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 3
LENGTH: 930
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-200-650E-3

Query Match 4.6%; Score 186.5; DB 4; Length 930;
Best Local Similarity 20.8%; Pred. No. 0.0006; Indels 199; Gaps 42;
Matches 175; Conservative 121; Mismatches 346;

QY 3 GEIASEKFNKLGNGKESLKK-DTGVHHNQ---ENESIKESKSFIDRNISTIRDFE 58
DB 56 GELNOSKNEITAPSNKTKTKKYVDSROLKDNQOTATADPKYTMDSATVKEKSSNMGSQ 115
QY 59 NKDLKLIKKKFREVDFTSETGRMBEYDYKYNDKNIIT-AYDDGTDELEYETKLDIK 117
DB 116 NATANQSTTKNSVNTTNKSSITVYSNE-----TDKSNLITQAKDVSTPKPTTIKPRTLN 169
QY 118 SKIYGVLSPEKDGHEILIGKISNVSKNAKVYVYNNVYSIEIKATYDHSKTMFTDLYAN 177

```

Db      170 RMAVNTVAAPQO-----TNV--NDKVHESN---IDIAIDKGVNQTGKTEFWAT 215
Qy      178 INDI-----VDGLAFAGDMRLFYKNDQKAKIKIMPEKIKETKSBYPVSSGVN 229
Db      216 SBDVLTALKANYTIDSVKEGD--TFYFKYGGYFRPESVRLP--SGTQNLVY--NAQGANI 267
Qy      230 IELGEGDLSKNRP-----DNLITMESGKIYSDSEKQOYLLKDNIIIRKGYALAKVTY 281
Db      268 IAKGIYDSTNTTNTTTFYFNVDQYTNVRG-----SFEQVAFAK-----RKATTDKTV 316
Qy      282 NPKGTDMLGNGVSKEDIKAKIOKANPNLRALSETTIYAD---SRNV-----EDGRST 331
Db      317 ---KHEVTLGNDYSEIIL--VDYGNKKAQPLISSYNNEDLSHMATVAVYQPNQVYT 371
Qy      332 QSVLMSALDGF-----NIIYQVTFPMNDKGEALIDKGLVYDSSGLVLFKGDDEKTYGE 387
Db      372 KQTFYTNLTGYKFNPAKNPKFI--YEVTKQNOQVD--SFTPDTSFL-----KQVLT-- 417
Qy      388 DKFNVEAIKEDGSMLEFIDTRPVNLSDKNYF-----NPSKSNKIYVRNBEFYLRGKI 439
Db      418 DQFDV-IYSNDMKTATVLDLKGQOTSNNQYIIQQVAVPNSSSTDN-----GKI 464
Qy      440 ----SDKGFNWE--LAVNESVDNLYLIYGDLDHIDTRDENIKLVNKG--DIMDKGM 489
Db      465 DYTLDTRTKYSMSNSYSNVNGSSTAN---GDQKKYNLGDYVWEDTNKDKODANEKGI 520
Qy      490 KQY-----KANGPP--DKYTDMDGNVYLQYSDLNKAVGVH-----YQFLYDNK 534
Db      521 KQYVYILKDSNGELEDRTTDENGKQFTGLSN-----GTYSVERSTPAGYPTTANVG 574
Qy      535 PEVNIDPKNTS--LEVADG-----KSVFNINDK--RNGPFGEIOEQHIYINCK 581
Db      575 TODAVDSGLTTTGVIKODNMNLTDSGFKTPKYSIGDYVWYDMSNDGKRDSTREKIKV 634
Qy      582 EYTSFNDIQIITDKTINIKIIVKDFARNTTVEFLNKOTGEVSEL--KPRHVTIION- 638
Db      635 KYTLQNEKEVEIGTT-----ETDENGKTRFNNLDSGKXVFEKRPAGILTQGTWT 684
Qy      639 -----GKMSSTVSEEDFILPYKGELEKGYQPDGMETSGEFGKKDAGVYINLSMD 690
Db      685 TEDDKDADGGEVDVTTTDDHF-----TLONGYEETSDSDSDSDSDSDSDSDSDSD 737
Qy      691 TEIKVPFKIEKEKEENKPTPDVSKKDNPCQVNSQLNESHKEDLQREHSQKSDYK 750
Db      738 S-----DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 788
Qy      751 D 751
Db      789 D 789

RESULT 14
US-09-200-650E-5
; Sequence 5, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eidiham, Delidre N1
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315

```

```

; TYPE: PR1
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-5

Query Match      4.6%; Score 186; DB 4; Length 1315;
Best Local Similarity 20.8%; Pred. No. 0.001;
Matches 188; Conservative 115; Mismatches 328; Indels 274; Gaps 51;

Qy      23 KQTTVEHHHNE-----ESIKKSSFTIDNISTIRPEKDL---KK 64
Db      164 KSVVNVQPTNEBKVDKAKTESITLNVKSDAKSDETLVDNN--SNSNNENADITLPPS 222
Qy      65 LKKKF-----REYVD--FTSETGRMEBYDYKDDKN--IAYDGDIDL 106
Db      223 TAPKLTNRMRILAAVQPSSTEAQVNDLITSTLTIVVD-----ADKNNKIYPADYLSL 277
Qy      107 EYETEKLEIKS-----KIYGLSPSKDGHFEILKISNVSKNAVYGGNNYK 154
Db      278 KQITVDDKVKSGDYFTIKYSDTVQVYG--LNDP-----DINKI--GDIKDPNNGE 324
Qy      155 SLEIKATKYDFSKMTF--DLVANINDIYDGLPA-----GDKRLPYK--DN 198
Db      325 T--LATAKHDIANMLITTFYTDYVDRNSVQMGINYSIYMDADTLPVSKNDVEFNVTIGN 382
Qy      199 DOKKAEIKIRMEKIKETKSEY-----PYSSYGN-----VIELGEGDLSKNK- 241
Db      383 TTTKTTANIQYPDYVNEKNSIGSAFTEYSHVGNKENGYKQTIYVNPSENSLTNAL 442
Qy      242 -----PNULTMESGKIYSD--SEKQOYLLKDNIIIRKQY-----ALKV 278
Db      443 KYQAVHSSYFNNI-----QINKDVTDIKIYQVPKGYTLNKGYDVWTKELTDTVNOYLQK 497
Qy      279 TTYNKGKTDMLGNGVSKEDIKAKIOKANPNLRALSETTIYADSRNVEDGRSTQSLMSA 338
Db      498 ITYGNNNSAVID---FGNADSAVVMWNTKQO-----YNS--BSPTLVQATLSS 543
Qy      339 LDGFNIIRQVQTFPMNDKGEA--IDKQNLV--TDSK-----LVLPCK 379
Db      544 TGNKVSSTGNALGFNNQSGGAGQGEYKIGNVWMBDTNNGQVBLEKGVNVTYVFN 603
Qy      380 DKEVTEGDKFVNEAIKEDGSMLEFT---DTRPVNLSDKNY--FNPSKSNKIYVRNBEF 433
Db      604 NNTTKYGE-----AVTKEDGSYLIPNLPGDYRVEFSNLPKGEYVTPSKQG---NNEEL 654
Qy      434 YLRKGISDKGCFNWEIARVNESVVDNLYLIGDLHIDTRDENIKLVNKGDIIMDKMKDYK 493
Db      655 -----DSNGLSSVITVNGK--DN--LSADLGI---YKPKNLV--GDYV--W--EDTN 693
Qy      494 ANGFPDK-----VTMDGNVYLQYGYSDLNKAVGVHGYFL--YDNVKEPVNI--D 540
Db      694 KNGIQDQDEKGISGVYTLKDENGNV--LKTVTTDADGK-----YKFTDLDNNGNYVEFTT 747
Qy      541 PKGNTSIEYADGKSVFNINDKRN---NGFDEGIOEHIYINGKEYTSFNDIKQIIDKT 596
Db      748 PEGYPTTYVTSDDIEKDSNGLTTTGVINGANMTLDSGFYKTPK---YNLGNVWMBDT 803
Qy      597 LNIKIVVDFANTYTKERILNKDGEVSELP--HRVYTIIONGKMSSTIYSEEDFILP 655
Db      804 -----NKDGKQDSTEGISGVYTLNKN-----ENGEVLD 832
Qy      656 VKGELEKGYQPDGMEISGF--EGKKDAGVYINLSKQTFIKVPFKIEKEKEENKPTPD 713
Db      833 TTKTKDQKYQFTGLENGYKVEFEPSPGYFT-----QVSGGTGDISNGSTTG 884
Qy      714 VSKKDNPOVNSQLNESHKEDLQREHSQKSDTKD-----VTATYLDKNMISSKST 767
Db      885 VIKDKNDTIDSGFYKFTYNLGDYVWEDTNKNGVODKDEKGISGVYTLKDENDKVLKTV 944
Qy      768 TNNPN 772
Db      945 TTDEN 949

```

```

RESULT 15
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PR1
; ORGANISM: Human
US-09-949-016-11433

```

```

Query Match 4.6%; Score 185; DB 4; Length 2733;
Best Local Similarity 20.0%; Pred. No. 0.0032;
Matches 170; Conservative 150; Mismatches 338; Indels 194; Gaps 39;

```

```

QY 21 LKDDTGVHHQENEE--SIKESFTIDRNIISTIDPENKDKLIKK---FRVVD 75
DB 1319 LKERIAGLEERKQKNEESQTLNENKNTLSQIST-KQSELMLQEVTKMNLNQIOE 1377
QY 76 FTSETGKMEEDYKYDD-----KGNIIAY-DDGTDLLEYETEKLEIKSIY 122
DB 1378 ELSRVTKLKETBEEDDLERLNLQALINGSIGVYCDVTDQIKNELLESEKMLK 1437
QY 123 VLSPEKDHFEILGKISNVSKAKVYY-----GN-----NYKSIETKA 160
DB 1438 CVSELEERKQOLVKEKTVSEIRKRYLEKIQAOKEPENKSHAKLOELKEOEVOQ 1497
QY 161 TKYD---FHSKTMTFDLVANDIYDGLAFAGDMLFVKNDDOKKAI-KIMPEKIKET 216
DB 1498 LQKDCIRIOEKISALERTYKALE-----FVOTESQKOLEITKENLAQAVEHR 1544
QY 217 KSEYPVSSYGNVIELEGEDLSKNRPDLTKMESGKIYSDSEKQOYLKDNIIIRK--- 272
DB 1545 KKAQAEIASFKVLLDDTQSEARVLADNLKKELOSNKESYKSGMKQKDBLERLEQA 1604
QY 273 -GYALKVTTYNPGKTDMLFGNGVYSKEDIQKANKPNLRALSETTIYADSRNVEDGRST 331
DB 1605 EEKHLKEKKMQEKDALAREKYLHEETIGEIQ-----VTLNKQKKEVOO--- 1649
QY 332 QSVLMSALDGFNIIRYQVTFPGKNDKGBAIDKGNLVTTDSKVLFGKDDKEYTGADKFN 391
DB 1650 ---LQENLDS-TVYQLAAFTKSM-----SLQDDRDRVIDEAK---KMERKFSDAIQSK 1696
QY 392 VEAI--KEDG-SMLEIDTKPVNLSDKNYFNSK---SNKIY--VRNPEFYLRGKISDK- 442
DB 1697 EEEIRLKEDNCSVLKQOLRQMSIHMEELKINISRLHDKQIWESKAQTEVQLOQKVCPTL 1756
QY 443 GGFNNEILAVNESVNNLYIGDLHDNTRDFNIKLVKDGDIIMDWGMKDYKANGFPDKYT 502
DB 1757 QGENKELL-----SOLEETR-----HLYHSQNELAKLESELKSLKDQLT 1796
QY 503 DMD-----GNV--YLQTYGSDL-NAKAVGVHYQFLYDNVKKREV----- 537
DB 1797 DLSNLEKCKEKGKGNLBGIIIRQOEADIONSK-----FSYQLETDLOASRELTSLRHE 1849
QY 538 NIDPKGNTSIEYADGK--SVVENINDKRNNGFGEIOEQHIYINGKEYTSF---NDIKOI 592
DB 1850 EIMNKEQKTIISLSGKEBAIQVAIAELRQO-HDKELKELENLSQEEENIYVLEENKKA 1908

```

```

QY 593 IDKTLNIKIVKDPARNTTVEFLINKDQGEVSEIK--PHRYVTIIONGKNSITIVSE 650
DB 1909 VDKTQIMLETIKTIKENIQOKAQDLSFVKSMSSIQNDRDRIVDPYQOLEERHLSILEK 1968
QY 651 DFLPVYGELEKGYQFDGWEISGPEG-----KKDAGVIVLSKOT 691
DB 1969 DOLIOEAALENKKKE-----EIRGLRSMDDLSNADLDAELIYREDLNOVITI-KDS 2023
QY 692 FIKPVFK-KIEEKEENKPTFDVSKKQDNQVNHQJNES-----HKREDLOREESHQK 745
DB 2024 QOKQLLEVQLOQNKLEKRYAKLEBKLESEBAN-EDLRSPNALQEEKQDLSKEIESIK 2082
QY 746 ---SDSTDVTA 754
DB 2083 VSIQLETRQVTA 2094

```

Search completed: February 16, 2005, 19:20:41
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:19:57 ; Search time 141 Seconds
(without alignments)
1791.327 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGEIAESKFKNLGNKESG.....ATVLDKNISSTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	773	US-10-067-385-8	Sequence 8, Appl1
2	4026	100.0	2119	US-09-769-744A-28	Sequence 28, Appl1
3	4026	100.0	2140	US-10-282-122A-73670	Sequence 73670, A
4	4026	100.0	2140	US-10-472-928-1180	Sequence 1180, Ap
5	615	15.3	117	US-09-765-872-68	Sequence 68, Appl1
6	227	5.6	1196	US-10-282-122A-52737	Sequence 52737, A
7	222	5.5	861	US-09-820-843A-109	Sequence 109, Appl1
8	218	5.4	1639	US-10-087-464-10	Sequence 10, Appl1
9	212	5.3	1141	US-10-282-122A-70251	Sequence 70251, A
10	207.5	5.2	4688	US-10-282-122A-76865	Sequence 76865, A
11	207	5.1	1191	US-10-282-122A-52048	Sequence 52048, A
12	203	5.0	1166	US-10-744-616-7	Sequence 7, Appl1
13	203	5.0	1881	US-10-032-585-7646	Sequence 7646, Ap

14	200.5	5.0	1786	9	US-09-742-096-3	Sequence 3, Appl1
15	200.5	5.0	1787	15	US-10-415-253-2	Sequence 2, Appl1
16	196.5	4.9	1184	15	US-10-282-122A-53254	Sequence 53254, A
17	196.5	4.9	5176	16	US-10-437-963-150986	Sequence 150986, A
18	195	4.8	1849	16	US-10-637-544-2	Sequence 2, Appl1
19	194	4.8	1178	15	US-10-282-122A-52434	Sequence 52434, A
20	193	4.8	1957	15	US-10-369-493-2070	Sequence 2070, A
21	193	4.8	903	15	US-10-282-122A-52328	Sequence 52328, A
22	193	4.8	909	15	US-10-282-122A-52109	Sequence 52109, A
23	192.5	4.8	708	17	US-10-472-928-3792	Sequence 3792, Ap
24	191.5	4.8	1156	15	US-10-369-493-43	Sequence 43, Appl1
25	190	4.7	1009	15	US-10-282-122A-43832	Sequence 43832, A
26	190	4.7	1788	15	US-10-282-122A-46664	Sequence 46664, A
27	189.5	4.7	841	9	US-09-815-242-5779	Sequence 5779, Ap
28	189.5	4.7	841	9	US-09-815-242-15751	Sequence 12751, A
29	189.5	4.7	1143	15	US-10-369-493-11081	Sequence 11081, A
30	189	4.7	1313	15	US-10-282-122A-76863	Sequence 76863, A
31	188	4.7	943	17	US-10-741-849-7309	Sequence 7309, Ap
32	187	4.6	1103	15	US-10-282-122A-76866	Sequence 76866, A
33	187	4.6	2265	15	US-10-282-122A-45123	Sequence 45123, A
34	186.5	4.6	930	17	US-10-744-616-3	Sequence 3, Appl1
35	186.5	4.6	1790	15	US-10-369-493-1586	Sequence 1586, Ap
36	186.5	4.6	1503	16	US-10-766-993-3	Sequence 3, Appl1
37	186	4.6	1315	17	US-10-744-616-5	Sequence 5, Appl1
38	186	4.6	1385	15	US-10-282-122A-44324	Sequence 44324, A
39	185	4.6	3325	16	US-10-408-765A-254	Sequence 254, App
40	185	4.6	5005	15	US-10-282-122A-76871	Sequence 76871, A
41	184.5	4.6	1946	15	US-10-282-122A-62947	Sequence 62947, A
42	184	4.6	1875	15	US-10-369-493-22285	Sequence 22285, A
43	183	4.5	2402	16	US-10-661-809-20	Sequence 20, Appl1
44	182.5	4.5	882	15	US-10-282-122A-53247	Sequence 53247, A
45	182.5	4.5	923	16	US-10-771-931-31	Sequence 31, Appl1

ALIGNMENTS

US-10-067-385-8	100.0%	Score 4026;	DB 13;	Length 773;
Best local Similarity	100.0%	Pred. 1.6e-216;		
Matches 773;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0;
Query Match				
1 KLGEIAESKFKNLGNKESGLKDDTTGVHHHQNENSIKESSTIDNNISTIRDPENK 60				
1 KLGEIAESKFKNLGNKESGLKDDTTGVHHHQNENSIKESSTIDNNISTIRDPENK 60				
61 DLKLIKKKFRVDPSTFTGRMEYDYKDYDKGIIIVYDGTDLLEYETKLDIKSKI 120				
61 DLKLIKKKFRVDPSTFTGRMEYDYKDYDKGIIIVYDGTDLLEYETKLDIKSKI 120				
61 DLKLIKKKFRVDPSTFTGRMEYDYKDYDKGIIIVYDGTDLLEYETKLDIKSKI 120				
121 YGVLSPKDGHFEIIGKISNVSKNAKVYGGNNYKSIKATKYDFHAKTMTFDLYANIND 180				

	Matches	773	Conservative	0	Mismatches	0	Indels	0	Gaps	0
Qy	1	KLGEIASKRNKINGKEGSLKXDTTGVHHHONNEESIKEKSSFTIDRNISITTRPEFNK								60
Db	1313	KLGEIASKRNKINGKEGSLKXDTTGVHHHONNEESIKEKSSFTIDRNISITTRPEFNK								1372
Qy	61	DLKLILKKKREVDPTFSEFGKMEEXDYKYDDKGNIIAYDDGTDLFEYFEKLDEIKSI								120
Db	1373	DLKLILKKKREVDPTFSEFGKMEEXDYKYDDKGNIIAYDDGTDLFEYFEKLDEIKSI								1432
Qy	121	YGVLSPEKQGHFELIGKISNVSKNAKYIYGNNTKSIITATKTIDFHSKTMTPDLNAIND								180
Db	1433	YGVLSPEKQGHFELIGKISNVSKNAKYIYGNNTKSIITATKTIDFHSKTMTPDLNAIND								1492
Qy	181	IYDGLAFAQDMRLFPVKNDQKAEIKIRMEPEKIKETSEKPYVSSYGNVIELGEGDLSKN								240
Db	1493	IYDGLAFAQDMRLFPVKNDQKAEIKIRMEPEKIKETSEKPYVSSYGNVIELGEGDLSKN								1552
Qy	241	KPDNLTMESGKTIYSDEKQOYLKONIILRKGYALKVTTYNPGKTDMLFGNGVYSKEDI								300
Db	1553	KPDNLTMESGKTIYSDEKQOYLKONIILRKGYALKVTTYNPGKTDMLFGNGVYSKEDI								1612
Qy	301	AKIQKAPNIRALSETTIYADSRNVEGRSTOSVLSALGPNIRIRQVTFPKKNDGEA								360
Db	1613	AKIQKAPNIRALSETTIYADSRNVEGRSTOSVLSALGPNIRIRQVTFPKKNDGEA								1672
Qy	361	IDXDGNLVTDSSKLVLPFGKDKKEYTGEBKFNVEAIKEDGSMFLDITKPVNLSMDKNYFNP								420
Db	1673	IDXDGNLVTDSSKLVLPFGKDKKEYTGEBKFNVEAIKEDGSMFLDITKPVNLSMDKNYFNP								1732
Qy	421	SKSKKTIYVRNPEFLPRKGISDKGFPNMLFVNESVDVNYLIYDGLHIDNRDPNLIKLVK								480
Db	1733	SKSKKTIYVRNPEFLPRKGISDKGFPNMLFVNESVDVNYLIYDGLHIDNRDPNLIKLVK								1792
Qy	481	DGDIMDMGMDYKANGPDPKYTMDGNYVLIQTGYSIDANAAVGHVQFLYDNNVKEPVNID								540
Db	1793	DGDIMDMGMDYKANGPDPKYTMDGNYVLIQTGYSIDANAAVGHVQFLYDNNVKEPVNID								1852
Qy	541	PKGNTSIEIYADGKSVVNIINDKRNNGPDEIIOEHIYINKEKTSFNDIQIIDLKTLINIK								600
Db	1853	PKGNTSIEIYADGKSVVNIINDKRNNGPDEIIOEHIYINKEKTSFNDIQIIDLKTLINIK								1912
Qy	601	IYVNDPFRNTTVKBFILNKDTGEVSELSKPRVTTYIONGKEMSTIYSEEDFILPVYKGE								660
Db	1913	IYVNDPFRNTTVKBFILNKDTGEVSELSKPRVTTYIONGKEMSTIYSEEDFILPVYKGE								1972
Qy	661	LEKGYQDFGWEISGFEGEKDAGVYINLSKDTFLIPVFKKIEKKKEENKCTPFDVSKKDN								720
Db	1973	LEKGYQDFGWEISGFEGEKDAGVYINLSKDTFLIPVFKKIEKKKEENKCTPFDVSKKDN								2032
Qy	721	PQVNHSQLNESHKREDIQREESHOKSPSTDVATVLDKNISISKSTNNPNK								773
Db	2033	PQVNHSQLNESHKREDIQREESHOKSPSTDVATVLDKNISISKSTNNPNK								2085
RESULT 3										
US-10-282-122A-73670										
Sequence 73670, Application US/10282122A										
Publication No. US20040029129A1										
GENERAL INFORMATION:										
APPLICANT: Wang, Liangsu										
APPLICANT: Zamudio, Carlos										
APPLICANT: Malone, Cheryl										
APPLICANT: Haselbeck, Robert										
APPLICANT: Ohlsen, Karl										
APPLICANT: Zyskind, Judith										
APPLICANT: Wall, Daniel										
APPLICANT: Trawick, John										
APPLICANT: Carr, Grant										

```
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patentin version 3.1
SEQ ID NO 73670
LENGTH: 2140
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670
```

```
Query Match 100.0%; Score 4026; DB 15; Length 2140;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KGEIASEKFKNLGNGKESGLKDDTTGVHHQNEESIKESSTFIDNISTIRDFENK 60
DB 1334 KGEIASEKFKNLGNGKESGLKDDTTGVHHQNEESIKESSTFIDNISTIRDFENK 1393
QY 61 DLKLIKKKFRVDDFTSETGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 120
DB 1394 DLKLIKKKFRVDDFTSETGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 1453
QY 121 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYSIEIKATKYDFHSKTMTPDLVANI 180
DB 1454 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYSIEIKATKYDFHSKTMTPDLVANI 1513
QY 181 IYDGLAFADDMRLFYVNDQKKAIEIKIRMEPIKETKSEYPVSSYGVNIELEGESLSKX 240
DB 1514 IYDGLAFADDMRLFYVNDQKKAIEIKIRMEPIKETKSEYPVSSYGVNIELEGESLSKX 1573
QY 241 KPDNLTMSGKTIYDSSEKQOYLKDNIIIRKGYALKVTTYNPGKTMLEGNVYSKEDI 300
DB 1574 KPDNLTMSGKTIYDSSEKQOYLKDNIIIRKGYALKVTTYNPGKTMLEGNVYSKEDI 1633
QY 301 AKIQKPNMLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPRMNDKGEA 360
DB 1634 AKIQKPNMLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPRMNDKGEA 1693
QY 361 IDKGNLVTDSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLPITKPVNLSMDKNYFNP 420
DB 1694 IDKGNLVTDSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLPITKPVNLSMDKNYFNP 1753
QY 421 SKSNKTYVNNPEYLLAGKISDKGFWELRVNESVVDNLIYGDHAIIDNTRDNIKLANKY 480
DB 1754 SKSNKTYVNNPEYLLAGKISDKGFWELRVNESVVDNLIYGDHAIIDNTRDNIKLANKY 1813
QY 481 DGDIMDMGKDYKANGFPDKVTMDGNVYLTQGYSDLANAKAVGVHIOPLYDNVKNPEVNI 540
DB 1814 DGDIMDMGKDYKANGFPDKVTMDGNVYLTQGYSDLANAKAVGVHIOPLYDNVKNPEVNI 1873
QY 541 PKGNISIEYADGKSVFENIDNRKNGFDEIOEHYINGKEYTSFNDIKQIIDTKLTANIK 600
```

```
DB 1874 PKGNISIEYADGKSVFENIDNRKNGFDEIOEHYINGKEYTSFNDIKQIIDTKLTANIK 1933
QY 601 IYVXDPARTVTKPEPLINDQGEVSELKPHRTVITONKEMSTTVSEEDTILPYKGE 660
DB 1934 IYVXDPARTVTKPEPLINDQGEVSELKPHRTVITONKEMSTTVSEEDTILPYKGE 1993
QY 661 LEKGVOFDGMEISGEGKKDAGVINLSKOTFIKPVFKKIEEKEENKPTFDVSKKON 720
DB 1994 LEKGVOFDGMEISGEGKKDAGVINLSKOTFIKPVFKKIEEKEENKPTFDVSKKON 2053
QY 721 POWNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSTXTNNPNK 773
DB 2054 POWNHSQLNESHKEDLOREHSQKSDSTKDVATVLDKNNISSTXTNNPNK 2106
```

RESULT 4

```
US-10-472-928-1180
Sequence 1180, Application US/10472928
Publication No. US20050020813A1
```

GENERAL INFORMATION:

APPLICANT: CHIRON SPA

TITLE OF INVENTION: THE INSTITUTE FOR GENOMIC RESEARCH

FILE REFERENCE: P026926WO

CURRENT APPLICATION NUMBER: US/10/472,928

PRIOR FILING DATE: 2003-09-26

PRIOR APPLICATION NUMBER: GB-0107658.7

NUMBER OF SEQ ID NOS: 4979

SOFTWARE: SeqMin99, version 1.03

SEQ ID NO 1180

LENGTH: 2140

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

FEATURES:

OTHER INFORMATION: serine protease, subtilase family

OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LpyTG)

OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)

US-10-472-928-1180

```
Query Match 100.0%; Score 4026; DB 17; Length 2140;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KGEIASEKFKNLGNGKESGLKDDTTGVHHQNEESIKESSTFIDNISTIRDFENK 60
DB 1334 KGEIASEKFKNLGNGKESGLKDDTTGVHHQNEESIKESSTFIDNISTIRDFENK 1393
QY 61 DLKLIKKKFRVDDFTSETGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 120
DB 1394 DLKLIKKKFRVDDFTSETGKRMEDYDYKDDKGNIIAYDGTDLLEYTEKLDKSKXI 1453
QY 121 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYSIEIKATKYDFHSKTMTPDLVANI 180
DB 1454 YGVLSPSKDGHEIILGKISNVSKNAKVYGNVYSIEIKATKYDFHSKTMTPDLVANI 1513
QY 181 IYDGLAFADDMRLFYVNDQKKAIEIKIRMEPIKETKSEYPVSSYGVNIELEGESLSKX 240
DB 1514 IYDGLAFADDMRLFYVNDQKKAIEIKIRMEPIKETKSEYPVSSYGVNIELEGESLSKX 1573
QY 241 KPDNLTMSGKTIYDSSEKQOYLKDNIIIRKGYALKVTTYNPGKTMLEGNVYSKEDI 300
DB 1574 KPDNLTMSGKTIYDSSEKQOYLKDNIIIRKGYALKVTTYNPGKTMLEGNVYSKEDI 1633
QY 301 AKIQKPNMLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPRMNDKGEA 360
DB 1634 AKIQKPNMLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPRMNDKGEA 1693
QY 361 IDKGNLVTDSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLPITKPVNLSMDKNYFNP 420
DB 1694 IDKGNLVTDSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLPITKPVNLSMDKNYFNP 1753
```

QY 421 SKSNKIYVBNPEFLRGKISDKGFMWELRVNESVVDNLYIGDHLIDNTRDNPNTKLVK 480
DB 1754 SSNNKIYVBNPEFLRGKISDKGFMWELRVNESVVDNLYIGDHLIDNTRDNPNTKLVK 1813
QY 481 DGDIMDMGKDYKANGFPDKVTMDMGNVYLQGYSPDLNKAAGVHYQFLYDVKKEVND 540
DB 1814 DGDIMDMGKDYKANGFPDKVTMDMGNVYLQGYSPDLNKAAGVHYQFLYDVKKEVND 1873
QY 541 PKGNTSIEYADGKSVFENINDKRNNGEDGIEIOHIIYINGKEYTSFNDIKQIIDKTLNK 600
DB 1874 PKGNTSIEYADGKSVFENINDKRNNGEDGIEIOHIIYINGKEYTSFNDIKQIIDKTLNK 1933
QY 601 IYVKDFARTTVEFLNKDGTSEVSELKPHRYVTYITQNKEMSTTVSEDFILPYKGE 660
DB 1934 IYVKDFARTTVEFLNKDGTSEVSELKPHRYVTYITQNKEMSTTVSEDFILPYKGE 1993
QY 661 LEKGQFDMGELSGEFGKADAGVYINLSKDTFIKPVFKKIEEKEBENKPTDVSCKKN 720
DB 1994 LEKGQFDMGELSGEFGKADAGVYINLSKDTFIKPVFKKIEEKEBENKPTDVSCKKN 2053
QY 721 POWNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 773
DB 2054 POWNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 2106

RESULT 5

US-09-765-272-68
Sequence No. 68, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match 15.3%; Score 615; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.4e-27;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGWSISGFBGKKDAGVYINLSKDTFIKPVFKKIEEKEBENKPTDVS 716
DB 1 YKGELEKGYQPDGWSISGFBGKKDAGVYINLSKDTFIKPVFKKIEEKEBENKPTDVS 60
QY 717 KCDNQVNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 773
DB 61 KCDNQVNHSQLNESHKEDLQREHHSQKSDSTKDYATATVLDKNNISSKSTNNPNK 117

RESULT 6

US-10-282-122A-52737
Sequence 52737, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 52737
LENGTH: 1196
TYPE: prt
ORGANISM: Clostridium botulinum
US-10-282-122A-52737

Query Match 5.6%; Score 227; DB 15; Length 1196;
Best Local Similarity 20.2%; Pred. No. 0.00029;
Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;

QY 1 KLGEIAESKFPKNGKSG-----SIKDDTGVHHHGHENESI 39
DB 123 RLKDIQIE-LFMDYTGIGKGGYSLIGGKIEAVLSGKPEERRSLLEBAAGIVKPKRKEAD 181
QY 40 KEKSFTLDNISTRDEN--KDIKKLIKKKFREVDV--TSETGKME----- 85
DB 182 KGLSN--TEQNLIRIKDILNTYEEEMPELKESEKAKKFLNLSBELKKEVNVVMYISDK 239
QY 86 -EYDKYDKNITIAVDGTD--LEYETEKLBELISKIYGVSPSDGHFELIGKISYNSK 143

```

Db      240 IEKDLK-NISSNMLSGENIDMLKNEKSOYEKIEISKF-----NEKLELDLK--NNSG 288
Qy      144 NAKVYGNPNKSEIKATKDYDFHSKMTEDPLVANINDIVDGLAFADGMLFVKDNDQCKA 203
Db      289 NKEEYTNKDKNKDIE-----NENALKEKIKNLDK-----NKEVENTLKTN 331
Qy      204 EIKIR-----MPEKIKETKSEYPPV-----223
Db      332 EERLQYVLEKKEALKEKINKLKEEESYLREIEIKGEDNVNPNKLEKEKELKIKSE 391
Qy      224 -----SYGNVYELGEGDL--SKNCPNLTMSGKISY-----DSKQOYLLK 265
Db      392 IELNSTNMLRNEISVMEWEIVLLENKLDNRKNSCDSYISSINIKETKEDIKEIKIK 451
Qy      266 DNILL-----RKGYALKVTTNPGKIDMLEGNQVSKED-----IAKIOK--AN 307
Db      452 ENLLENNLKNERSNISLKSISLNNKEK-KLKEKNAAYSRLAANYHMLSNLEKHYEGYN 510
Qy      308 PNLRALSETTIYADSRNVEDGSRSTQSVLMSALDGFNIIRYQVTFPKMDKGAIDKXGML 367
Db      511 RSVKLTMEHVSYGKVNINIGG-----GEVLGDIIKVKKELETAMEIALGAI-----SNV 560
Qy      368 VYDSKVLVFGKDDKEYTEGDKFNVEAIKEDGSMLEFIDTKPVNLSDMKYFNP--SKSN 424
Db      561 IYEDD-----NKAKILINYLKCK-----SLGRATFLPLTTIOGR 594
Qy      425 KIYVRNPEFYLRKGISDKGSGFMNELRVNESVVDNYLIYGDHIDNTRDPNI--KLNVKOG 482
Db      595 KAKIN-----VTRDEGF--LGIASDLIY--DYKFSNIIDIVLGRTLVAKD- 637
Qy      483 DIMDMGMDKYKANGFPDKYTDMDGNVYLTQYGSDLNAKAVGVHGYOFLYDNVKEPVNIDPK 542
Db      638 --MDSALKLAKLNTYSPFKIYTLLEGV-----INPG 665
Qy      543 GNT--SIEYADKSVVFNINDKRNNGFDGEIOEHOIYINGKEYTSFNDIKIIDTLNLI 599
Db      666 GSIYLGSGIKHRASSI--ISRRK-----EIRE-----TKKLEBETKN- 700
Qy      600 KIYVDFANTYKKEPILNKDGTGEVSELKPHRTVITLIONGKEMSTIVSEEDLITLVYK 659
Db      701 -----TTEEFNKN-----ILENNKIKIT--LDEENLNT--KD 728
Qy      660 ELEKGYOPDGEWISGFEKKDA-----GYVINLSKD--TFIKVFKKIEKEKEENK 709
Db      729 EI-----YANNIELTKMGKMLAIKEDTERLRSSLSINISREIKLTKOKIOEENINISQK 784
Qy      710 PTFDVSKKQDPVNSQNLNESHKEDLQREHSHQSDSKY--DVTATVLDKNNISK 765
Db      785 QLEELKLKCD--LNNHNDICE--EDFLQNEEENVKNIKDKLIEYKIEKAKLDEMVLVSIK 839
Qy      766 -----STNNPNK 773
Db      840 KELYSMDTNTITNLNENK 857

RESULT 7
US-09-820-843A-109
; Sequence 109, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820, 843A
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 109
; LENGTH: 861
; TYPE: PR
; ORGANISM: Plasmodium falciparum
; FEATURE:

```

```

; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|4493994
US-09-820-843A-109

Query Match      5.5%; Score 222; DB 10; Length 861;
Best local Similarity 19.3%; Pred. No. 0.00037;
Matches 162; Conservative 136; Mismatches 324; Indels 216; Gaps 37;

Qy      2 LGEIASEKPKNTANG-----KESGLKQDTTGYEHHH--OEENESIKESKSFIT 47
Db      154 LNNINGKVFKNIDDDCNVPLPTNNLYIDKEG--KMLTGEKHYNAASNEYNHNNKNTNYY 211
Qy      48 DNISTIRDFEKKDL-----KKLKKKREVDDFSFGKMEEDYKDYKDDGNTI 98
Db      212 NNNSTNNNNPFCNNNYNDNNYNNNSNNGKNGKTERSLNLYK--KEHDVYDEYENNGNTR 268
Qy      99 AYD-----DGTDLVEYTEKLDEIKSKIYGVLSPSKDHFEI--LQKISNVSRNAKYVYN 151
Db      269 KNDSEKYNWNPPLHY-----SKNNVYDIFLGDIKKAYAKNNEKKN 310
Qy      152 NTKSIEIKATKYDFHSKMTFPLVANINDIVG--LAFAGMRLFVKDNDQCKAEIKIR 208
Db      311 KTNMNM-----HDNNSN--NSNNVNLNNNNNSNNYNNIFKONDENTL----- 351
Qy      209 MEKIKETKSEYVYSYGVNVEIGEG-DLSK--NKPNLTMSGKLYSPSEKQOYLL 264
Db      352 -----TNSNPAKPFKNNNNNVNEENTDIKYLANK-----NSQHSIDCKNNNNN 396
Qy      265 KNIILRKGYALKVTTNPGKIDMLEGNQVSKEDIKIQKANPN--LRALSETTIYAD 321
Db      397 GNNII-----NNSNNKNNIPQNGSRNRYENWVYNNNNNNNNIISNNKREAFNPD 447
Qy      322 SRNVEDGSRSTQSVLMSALDGFNIIRYQVTFPKMDG-BAIDKQNLVYDSKVLVFGD 380
Db      448 NINTNSGREEEKISNTVAE--LLMKQISMIKERNGLDVLREKNT-----FGFL 494
Qy      381 DKEYTEGDFNVEAIKEDGSMLEFIDTKPVNLSDMKYFNPSKNTKYVRNPEFYL----- 435
Db      495 DNNYQNYGSNN-----NSSLEKNNM--KENDIYSKASGRINDIR 533
Qy      436 ---RGKISDKGSGFMNELRVNESVVDNYLIYGDHIDNTRDPNIKLVNKGDIIMDMGMD 491
Db      534 TINSNGVLSQF-----SLVNGSVLNNNNNNYNNNSNRRNKNQNNNNNNNNNNM----- 582
Qy      492 YKANGFPDKYTDMDGNVYLTQYGSDLNAK--AVGYHYQFLYDNVKEPVNIDPKNTSIE 548
Db      583 -----NNNNNSNNNNNNNNNNNNYNNNNHKKYHSMQDVTYKIF--INNYGNNNGNNNSNS 633
Qy      549 YADGKSVFNINDKRNNGFDGEIOEHOIYINGKEYTSFNDIKIIDK-----595
Db      634 NSNNNVYEHYYNNKKN--FKAKLNNTYHMLPDKNKNNMANNNTYNNIKNNLSMNEPNPPL 691
Qy      596 TLNIKIVVDFAFN--TTVKEFILNKDGTGEVSELKPHRTVITLIONGKEMSTIVSEED 651
Db      692 SFNNSDINKNNQGNINITPIINSILRLD-NEVDVHNNSISEBIGNAK--VSNVLDLSK 748
Qy      652 FILPVYKGLBKGYQDGEWISGFEKKDAQGVYINLSKTPFK-----PVFKIEKK 704
Db      749 SLTKSKSGGNNNNYNP--KNFNNNNN-----NNNSKSFITNYSQOYYPSHQOQOQH 799
Qy      705 EENKKTPTVSKKQDPVNSQNLNESHKEDLQREHSHQSDSKY--SDSTQVATVLDK 759
Db      800 QQQQQQ--QQQQTLOTQINSLHNDFNKKKPKRKYPMKTPBEFGTTNEMVAK 855

RESULT 8
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishli, Athar

```

```

; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: SI237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match      5.4%; Score 218; DB 14; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.0014;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

QY 23 KDTTG-VEHHQENESIK-----EKSFTIDRNISTRIDPENKDLKLIKRRBYVD 75
DB 232 KDNVGMEDYIKAKKTIENINELIBESKTIIDKKNATKEE-----KKKLYQAOY 283
QY 76 FTSETGKMEEDYKYDKGNIAYDGTDLFETEKLDIKSIYGVLSPEKDGHFELL 135
DB 284 DLSYNNQOE-----AHNLISV-----LE-KRIDTLK-----KNEIKELL 319
QY 136 GKISVSKAKAKYVYNNYKSIKATYDFH-----SKMTFPLYNINIDVGLAFA 188
DB 330 DKINEIKPPPPNNTPTLADKNKKIIEHKEIKELAKTIKFNIDSLFTDPL----- 373
QY 189 GDMRLFVNDNDOKKAEIKIRMEKIKETSEYV-----YVSYGVN-----TELGE-GDLSK 239
DB 374 -ELEYYLKEKN-KATIDISAKVETKSTEPNEIPNGVTTPLSTINDINNALNELNSGDLI- 430
QY 240 NKPDMLYKMGSKYISDSEKQOYL--LKDNILL-----RKGYALKVTTYNPKRTDML 289
DB 431 -NPDYTKPEPSKNITVDNERKKFINIEKIKIEKKKIESDSEKSLN-----DIT 485
QY 230 EG-----NGVYSKE-----DIAKIOKA-----NPNRLSETTYIADSANVEDGSTOSV 334
DB 486 KEYEKLNEIYDSKNNNIDLTNFEKMGKRYSYVEKLTHTNTPASVENSFNLKLT 545
QY 335 IMSALDGF--NI-----RYQVFTPKNDKGEAIDKGNLVTSKLVLFSGDKDEXT 385
DB 546 ALKWMEDYSLRIVVEKELKYYKNIISKIENIEITLVE--NIKDEBQLF-----EKKIT 598
QY 386 GEDKFNVEAIKEDGSMFLIDTRPVNLSDMKYFNPSKSNKIYVRNPEFYLRGKISDKGPF 445
DB 599 KDNKPEDEKILEVSDIVKQVQKVL--MNKIDELKQTQLLKAVE--LKNHIVPNYSY 653
QY 446 NWEELAVNESVDNNTLIYGDHLIDNTRDRIK-----NVK----- 480
DB 654 KOENKOE--PYLILVILKEIDKLKVPKVESLINEKKNIKTEGOSDNSEPTSEGEI 709
QY 481 -----DGDIMDMGMDYK-----ANGFPDVK 501
DB 710 TGOATTKGQQAAGSALBEGDSVQAQAQEQQAOPPVVPEAKAQVPPPPAVNNKTEV 769
QY 502 TDMGNAVYIKQGYSDLANA-----KAVGYHYQFLYNNVKEPVNIDPKGNTSIEYADGKSV 556
DB 770 SKLD-----YLEKLYEFLNTSYICHKYLIVSHSTMEKILKQYKITEESKSSCDPLDL 826
QY 557 FNIN-----DKRNNGPDG--EIOEOHIIYNGKEYTSFNDIKQIIDLTLNLIKIV 603
DB 827 FNIONNIPWYSMFPSLNNLSQLFMEIYKEMVONLKYKDNNDIKMLLEBAKVSISV 886
QY 604 KOFANNTYKFEILNKDTGEVSELKPHRYTATYIONGKEMSTIYSEBDPILFVYGELEK 663
DB 887 KTLSSSNQPLSLTQDKREYVANDTSHSTNLSKLFENILS-----LQKNKN 937

```

```

QY 664 GYQPDGWEISGEFGK-----KDAGVINLSKDTPIKVPFKKIEEKEENKPTF--D 713
DB 938 IYQ-----ELIGQKSENFEYKILKSDPTFVNESFTNFKSKADINSINDSKRKLSED 993
QY 714 VSKKDNQNVNHSQNSHREKEDLOREHSQKSDSTKVATVLDKNNISK-STTNMP 771
DB 994 INKLKKTQLQSFDLNKKYKLTLERLFDPKKTGVGKYKQIKLTLTLKEQLSEKSLNLP 1052

RESULT 9
US-10-282-122A-70251
; Sequence 70251, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haebelbeck, Robert
; APPLICANT: Olesen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70251
; LENGTH: 1141
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-70251

Query Match      5.3%; Score 212; DB 15; Length 1141;
Best Local Similarity 20.9%; Pred. No. 0.0019;
Matches 182; Conservative 116; Mismatches 336; Indels 238; Gaps 44;

QY 33 GENESISE--KSFTIDRNISTRIDPENKDLKLIKRRF-----EYDDPTS 78
DB 221 KNPEKLELYVNDNNTDRSTKPVATFTVAAPKRLNKKMFPAVAQPAVASNNVNDLIT 280
QY 79 ETGKMEEDYKYDKGNIAYDGTDLFETEK-----KLDEIKSIKYGVLS 125
DB 281 VTKQIK-----VGGKDKVAAAHQKQDLEVDTEFTIDNKVKKGDMTINTNYDKNVPISLT 336
QY 126 PSKD-----GHFEILGKISNVSKNK--VYGNKYKSIKATYDFHSHKTMTPDL 174

```

[illegible]

```

Qy 522 VGHVHGYFLVDNVPKEVNIPEKNTSLEYADGSAVFNNDK-----RNNGFDEIEQH 575
Db 624 IDKRDILNLDG--HBPPTYNDGNLEIH-----TQLANDLNDLQKALNNANVAKIIVDQ- 677
Qy 576 IYNGKEYSFNDIKOILDKTLANIKIYVKDFPANTVYS--FILKDTGSEVSELKPHRV 633
Db 678 ---DGIEH-----EIDVSDANGKAVIIPYNNLANNPTPTNITYLTK----- 716
Qy 634 VTIQNKEMSSSTIVSEEE-----DFILPVYKGELEKGYOPDGWEIS----- 673
Db 717 VWLKNNQNPIDLISEQLSGDNHISFKKPTITAKTEN--DDYEISFNSPLANKIK 773
Qy 674 -----GFEKGDAGVINLSDPTIKGVFKVFKIEKKEEENKPFEDV 714
Db 774 LTFKTDNDNTTKTVEASIGLDGKA-----IFETSDAIFAPDHKTYTLTKLEADNKKVANI 829
Qy 715 SKKKDPVYNHSQLNESHKEDLOREHESOKSDSTKIDTATATVLNKKN----- 761
Db 830 DEISPLDRIYNKQKNQNVADN--KHFEKIPQOKKDLTAYVKKKNNEIHVPKTDKG 887
Qy 762 ---ISKSTNNPNK 773
Db 888 KXIVNPNNNLPEDNK 902

```

```

1 RESULT 11
2 US-10-282-122A-52048
3 Sequence 52048, Application US/10282122A
4 Publication No. US20040029129A1
5 GENERAL INFORMATION:
6 APPLICANT: Wang, Liangsu
7 APPLICANT: Zamudio, Carlos
8 APPLICANT: Malone, Cheryl
9 APPLICANT: Haselbeck, Robert
10 APPLICANT: Ohlsen, Kari
11 APPLICANT: Zybkind, Judith
12 APPLICANT: Wall, Daniel
13 APPLICANT: Trawick, John
14 APPLICANT: Carr, Grant
15 APPLICANT: Yamamoto, Robert
16 APPLICANT: Forsyth, R.
17 APPLICANT: Xu, H.
18 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
19 FILE REFERENCE: BLI/TPA.034A
20 CURRENT APPLICATION NUMBER: US/10/282,122A
21 CURRENT FILING DATE: 2003-02-20
22 PRIOR APPLICATION NUMBER: 60/191,078
23 PRIOR FILING DATE: 2000-03-21
24 PRIOR APPLICATION NUMBER: 60/206,848
25 PRIOR FILING DATE: 2000-05-23
26 PRIOR APPLICATION NUMBER: 60/207,727
27 PRIOR FILING DATE: 2000-05-26
28 PRIOR APPLICATION NUMBER: 60/230,335
29 PRIOR FILING DATE: 2000-09-06
30 PRIOR APPLICATION NUMBER: 60/230,347
31 PRIOR FILING DATE: 2000-09-09
32 PRIOR APPLICATION NUMBER: 60/242,578
33 PRIOR FILING DATE: 2000-10-23
34 PRIOR APPLICATION NUMBER: 60/253,625
35 PRIOR FILING DATE: 2000-11-27
36 PRIOR APPLICATION NUMBER: 60/257,931
37 PRIOR FILING DATE: 2000-12-22
38 PRIOR APPLICATION NUMBER: 60/267,636
39 PRIOR FILING DATE: 2001-02-09
40 PRIOR APPLICATION NUMBER: 60/269,308
41 PRIOR FILING DATE: 2001-02-16
42 Remaining Prior Application data removed - See File Wrapper or PAM.
43 NUMBER OF SEQ ID NOS: 78614
44 SOFTWARE: PatentIn version 3.1
45 SEQ ID NO 52048
46 LENGTH: 1191
47 TYPE: PRT
48 ORGANISM: Clostridium acetobutylicum

```

US-10-282-122A-52048

Query Match	5.1%	Score 207;	DB 15;	Length 1191;
Best Local Similarity	19.6%;	Pred. No. 0.0038;		
Matches 178;	Conservative 139;	Mismatches 275;	Indels 314;	Gaps 43

QY	1	KLGSJASCKFNLNGKGG-----	SLKQTTVEHNNHONBESSI	39
QY	120	RLKDIOE-LFMDTGLGKGGYSII	QOGKIEAVLSGKPERRRALLLEAAOIVUKFKTKYDA-	177
QY	40	KEKSFITDRNISITRD-----	FENKDLKLIKKKFFREVD--FTSETG--RMEE	86
Db	178	-EKLJENTNQVLVINDLITRYEERLEPLRIE	SEKAKFFVELDELTKYEINTIIVISIDN	236
QY	87	YDYKTYDDKGNIIA-----YDDGT-----	DLEFETKDEI-----KSKIYGLSP	126
Db	237	IDYRINDLKQMAULKJISIDENVKDEKIS	LELVATSELDLDFAKYSSNKTYY-----E	292
QY	127	SKOGHFEILIGKI-----SNVSKNAYYGN	NNYSKISIE-----IKATKYDFHSKTMTPDLY	175
Db	293	SKSEHQKLSIELLEKETSNSDPAKLYE	IELDLSINYLK-SRKEIQLKTIED--	349
QY	176	ANINDIVDGLAFAGDMRLFYVNDOKA	EIKIRMP--KIKETKSE---YPVYSSYGN	228
Db	350	KNYNK-----ELLSKINSSEKKNI	DGLLEWEMKSIQYONDADIIISTISQNN	400
QY	229	VIELGEBDLSNKNKDNLTYMES-----	GKIYDSEKQOVLKDNITLRGVALKVTYTPG	284
Db	401	EVVI-----LKKIEISNESKLESIR	PAGGYEKS-----LKINETYKTLSEELVKIN--	448
QY	285	KTMDLEGNVGYSKEDIARIKANPMLRL	SETTIYADSRNVEDGRSTQVLSMSALVDENI	344
Db	449	--DKISGYENQIRERNRSKISTLN--	RIISBE-----	475
QY	345	IRYQVFTFKMNDKGEAIDKDNLV	YDSSKVLFGDKDEYTGEDKFNVAIKEDGSMLEFI	404
Db	476	---EKLARELNSKSNKLEANKMLINL-	-----EKQEGYNR-SVKNLMQHTKGFV	522
QY	405	DTKRVNLSMDKNFNPNPSKSKIYR	ANPEFYRGKISDKGGRNWEKLRVNESVVD--	461
Db	523	DVKPESSFVLGEVIRKVKKEFETA	V---SISLGAISD-----ITLDDNIAKKLINYLK	573
QY	462	YGDJHINDTRDENI-----KLVKOG-	-----	482
Db	574	SKNLGRATFLPLNIIGKRLNIS	DPATRHEKFGITIASLEIDYDSTFLPAVNYLGRYIV	633
QY	483	DIMDMGMDKYKANGFPDKVTMDGN	VYLTQTSYDLNAKAVGVHYOFLYDNVAKPEVNI	542
Db	634	DNMDSALKIATLNGYSFKITVLTE	EV-----	663
QY	543	GNTS--IEYADGKSVFENINDKRN	NGPDEGIEOBHITYNGKBYEYSFNDIKQIIDTLN	600
Db	664	GSLTGSGSTYSKASII-----	GRKREIELEINLEINNYSQA---LEOSSNKIENK	710
QY	601	IYVNDPANTTVKPFILINKOGEV	SELKPHRVUTYIONGKSSMSTIVSEEDPLVYNGE	660
Db	711	KVAYE-----LDNLCD-----	LITTI-HEKIELETKIERLKSIDISEK	750
QY	661	LEKGYOFDGEWIEISFEGEKD	AGYVINYLSKDTFYIKVPFKIEEK--KEENKPTEDVSK	717
Db	751	LKNSYNTSVSGIG-----	FIEKIKNHLEKLVKVEENKA---LKLK	789
QY	718	KNDPQVNSQLNESHK-----	EDLOREHSGQSD-----STYDVAFTVLDK	759
Db	790	EAN---NNNLIDELERKLO	ENSKVLINEIBEMKVDKAKSDEMLMSSTREIEREYKEM	846
QY	760	NNISSK	765	
Db	847	HNMEK	852	

RESULT 12
US-10-744-616-7

/ Sequence 7, Application US/10744616
/ Publication No. US20050026170A1
/ GENERAL INFORMATION:
/ APPLICANT: Patti, Joseph M.
/ APPLICANT: Foster, Timothy J.
/ APPLICANT: Hook, Magnus A.O.
/ APPLICANT: Eidiham, Delidre N.I.
/ APPLICANT: Perkins, Samuel L.
/ TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
/ FILE REFERENCE: P06283US2/BAS
/ CURRENT APPLICATION NUMBER: US/10/744,616
/ PRIOR FILING DATE: 2003-12-24
/ PRIOR APPLICATION NUMBER: 60/066,815
/ PRIOR FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: 60/098,427
/ PRIOR FILING DATE: 1998-08-31
/ NUMBER OF SEQ ID NOS: 23
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 7
/ LENGTH: 1166
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-744-616-7

Query Match 5.0%; Score 203; DB 17; Length 1166;
Best Local Similarity 21.6%; Pred. No. 0.0062;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

33 QENESIRK-KSFTIDRNIITRDPENKDLKIKKFR-----EVDFFS 78
226 KNPPELKELVNDSNTDSTKVAATPSTVAPKRVNAKPAVAPAAVASNNVNDLIK 285
79 ETGKRMEEDYKVDKGNIIAYDDGTDEYETE-----KLDEIKSIKIVGS 125
286 VTKQITK---VGDGKNVAAAHADGKQIDIEFTIDNKKYKGDITMTIKVINDKIVSDLT 341
126 PSKD-----GHEPILKISNVSNAK---VYGNVNSYIEIKATYDFHSK----- 168
342 DKNDPIDITDPGEVIAKGPDKATQIITYTFDYDKEDISRLTIVSYIDKIVPNE 401
169 ---TMTF---DLVANN-DIVDGLAFAGD---MELPYK-DNDQKAEIKRM-PEKIK 214
402 TSLNLFATAGKETSQNVTVDYODPMVH-GDSNIQSIPTKLEDBKQITIQIIVNLIKKS 460
215 ETKSEY---PYVSSYGVNIELEGEED--LSKNKPDULTK-----ESGKIYSEKQ 261
461 AATTKVDIGSOVDYDIGN-IKLGNSFTIIDQTEIRIVKNSDQOLPQSRITVDSQYED 519
262 YLLK-DNIIIRKGYALKVTTYPNGKTDMLGNGVSGKEDIKIQKANPILRALSETTIVA 320
520 VTSQFNN---KKSFSNNVATIDRG-----DINSAYI-----IKVSKYTPPS 558
321 DSR-NVEDGRSTQSVLMSALDGFNIIIRYOVFTFRKNDKG-----E 359
559 DELDIAQGTSMKTT--DKYGYNNVAGYSNFIYTSNDTGGDGTVPBEKLYKIGDYVWE 616
360 AIDKQGNLVTDSKYLFGKDDKEYTGBKFNVEAIKEDGSMFLDTKPYVNSMDKNY-F 418
617 DVDKDD-----VQGTDSKEKPMANVL-VTLTYPDGT-----TKSRFDANGHYER 660
419 NPSKSNKIVV---RNBEFYLRGKIS-----DKGGSFNNELRVNESVNVNLYIGDLHD 468
661 GGLKQDEYTVKFEPTGYLPTKVNNGTTDGEKQNSSSVTVKINGK-----DMSJD 712
469 -----NTRDFNIIKANVKG--DIMWGMKDYRANGEPDKVITMDGNVYLQIGYSD 517
713 TSEYKPKNLDGYVWEDNKGIDANBPGIDVAVT-----LKDSTGCV-IGITTTDA 766
518 NAKAVGVHQLYDANKPEVINIDPKGNTSIEY---ADGKSVFN--INRKNNGPFGELQ 572
767 SGK---YKF-----TDLD-NQNYVEFEETPGYPTVYNTTADDSQGL----- 807
573 EGHIIYNGKEYTSFNDIKQIIDKTIKIVVDPAANT--TYKEPI---LNKDTGEVSEL 627

DB 808 -----TTTGVIKQADNMWTLID-----RGFYKPKYSIGADYVWDSNKGQKDSYE 851
628 KHR-VTYTIOGK-EMSTIYSEDDFILPVKSGLEKYOQDMEISFEKKGAGYVI 685
852 KQIKQVTVTLQNEKEVIGTITIDN-----GK---YRPD----- 883
686 NLSKQTFIKPVFKTIEKKEENKPTFDVSKKQNPV-----NHSQ--LINESHRKEDIQ 738
DB 884 NDSGKY-KVIFEKRAGLQTYVTNTED-DKQADGGEVVTITDHDFTLNDGFFEDTS 941
739 REHSQKSDSTXD 751
DB 942 DSDSDSDSDSDSD 954

RESULT 13
US-10-032-585-7646
/ Sequence 7646, Application US/10032585
/ Publication No. US20030180953A1
/ GENERAL INFORMATION:
/ APPLICANT: Terry, Roemer D.
/ APPLICANT: Bo, Jiang
/ APPLICANT: Charles, Boone
/ APPLICANT: Howard, Busesey
/ TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
/ FILE REFERENCE: 10182-005-999
/ CURRENT APPLICATION NUMBER: US/10/032,585
/ PRIOR FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 8000
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 7646
/ LENGTH: 1881
/ TYPE: PRT
/ ORGANISM: Candida albicans
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (1881)..(1881)
/ OTHER INFORMATION: X=any amino acid
US-10-032-585-7646

Query Match 5.0%; Score 203; DB 14; Length 1881;
Best Local Similarity 20.8%; Pred. No. 0.011;
Matches 172; Conservative 132; Mismatches 300; Indels 222; Gaps 39;

22 KQDTGV---HHQENESIRKESFTIDN-ISTIRDPENKDLKIKKGFREVDP 76
1140 KATDEIEKTKHITDLOEBAKQKQSFSEHNDIKSNLDEANKELSD-NREKLSNLEKE 1198
77 TSETGRMEEDYKVDKGNIIAY--DDGTDEYETEKLDEIKSIKIVGLSPSKDGFPI 134
1199 KTELANKLKTQBEKISDLETSVAISEDKSKLKHQIDELKRIKILUETTLKNEBEWTFK 1258
135 LCKISVNSKNAKVVYGNVNSYIEIKATKYDFHSKTMTPDLVANNIDVGLAFAGDMLF 194
1259 KEQLOVNDKCK-----ELBAC-----LKKLTKEKEINDLIKLEAA----- 1297
195 VQNDQKRAEIKIRPEKIKETKSEYPYVSSYGVNIELEGEGLSKKKPNLTKMESGKIY 254
1298 KSDHPTERRKSL---LIEDTKSR---SEKNVIKLE-----QIEKIK 1334
255 SDSEKQOYLLKNIIIRKGYALKVTTYPNGKTD-----LENGVSGKEDIKIQK 305
1335 GERKE---VRD---IQSLAKTIDWEIKITTLKDKSDELETT--NKESVDTLLK 1385
306 ANPNLR--ALSETTIYADSRNVEDGRSTQSVLMSALDGFNI-----IRYQVF- 350
1386 EVENLKEKISLEDDKQDQTKYKELAAQLFTKSNLSDSTWELKTELELKKNVRELTE 1445
351 ---TFKNDKG---EALDKQGNLVTDSKYLFGKDDKEYTGBKFNVEAIKEDGSMFL 402
1446 ATSEITKLODNNQSLTEIEKTKAALTVSSK-----DLEVCQGNQSELQ-----DSLK 1493

LENGTH: 1787
TYPE: PR
ORGANISM: K1 Parasite Clone
US-10-415-253-2

Query Match 5.0%; Score 200.5; DB 15; Length 1787;
Best Local Similarity 20.5%; Pred. No. 0.015;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AES--KPKK-LGNGKESLAKDT-----TGVEHHQENE-----ESIK 40
DB 868 AESVTTFSNILEIQENTITNDTIEBKLEHENVLSAALENTQSEKKEVIDVIEBVK 927
QY 41 EKSFTI-----DRNISTIRD--PEN-----KOLKLIKKEFREVDFTS 78
DB 928 EBAVATLLETVAQAEKSNATITTEIPENLENAVESNENVAENLEKLNETVFNVTLDKYE 987
QY 79 ET---GKME--EYDYKY-----DDKG--NIIAYDDGTDLLEYT-----EKLD-- 114
DB 988 ETVEISGSLENNEMDKAFPSFIDNVKGIQENLLT--GMFRSIFTSIVIQSEKVDLN 1044
QY 115 -EIKSKIYGVLSPSKDGHEIIGKISNVKNAKV-----YGNKYKSEIKATKYDFH 166
DB 1045 ENNVSSILDNINMKEG---LNLKLENISSTEGVQETVTEHVEQNVYVDVDPAMK---- 1097
QY 167 SKTMTFDLYANINDYDGLAFAGDMRLFYKNDQKAEIKIRNPEKIKETKSEYPVVSSY 226
DB 1098 -----DQFLGILNEAGL-----KEMFNLBDVPKSESDVITVEIKDPEVQKEVERET 1146
QY 227 GNVIELGEGD---LSKNRPDLTYK---ESGKIYSDSEKQOYLKDNIIIRKGYALK 277
DB 1147 VSIIEMEMENIYDLVEBEKEDLTDKWDIVABESSIEISDSKESTESIKDK--EKDVSIV 1203
QY 278 VTTYNG-----KTMLEGNNGVYSK--EDIAKIQKANPVL----- 310
DB 1204 VEEVDNDMDSEVEKYELKKNBEBELMKDAVEINDITSKLIEBTQELNEVADLIDMKER 1263
QY 311 -----RALSETTIYADSRNVEDOR--STQVLSALDGNFIIRYQVTFKANDKGEAIDK 363
DB 1264 LKELEKALSE-----DSKEIIDAKDPTLEKVIIEBHD-----ITTLDBEVVELKDV 1309
QY 364 DGNLVYDSSKLYLFGKDKKEYTGEDKFVNEBAIKEDGSMLEFIDTKPVNL-----SM 413
DB 1310 EEDKIEKVS DL---KDLIEDILKE---VKEIKELSEILBEDYKELKTIETDILEBEKEI 1362
QY 414 DKQYFNPSSKSNKIYVANKPEFYLRGKISDKGFGNMLELVNESVVD-----NYLIYDGL 465
DB 1363 EKDHFEKFEFEAEBEIKDLADILKEVS-----SLEVEBEKQLEVEHLEKEVEHIIISGDA 1417
QY 466 HLDNTRDFNIXLVKND--GDIM-----DMGMDKYKANGFPDKVTMDGNNVYLQTVGSDLN 518
DB 1418 HIKGLEEDDLF--EVDDLKSGIILDMKGDHGLGMDKESLEDDVTTKIGERY--ESLKDYL 1473
QY 519 AKAVGVHYQFLYDN-----VKPEVNIDPKGNTSIEYADGKSVFVNINDK--RN 564
DB 1474 SSALGMDDEQOMTKRKAQPKLEEVLLKEVEKEEPPKKIT-----KKKVRFDIKDKERPO 1528
QY 565 NGFDGIEIQOHYIYINK-----EYTSRNDIKQIIDKTLN-----IKIYKDPAR--N 609
DB 1529 EIVEVEMKDEDIEEDVEBDEIEDIEDKVEDIDEDIDEDIDEDKDEVIDLIYQKEKRIEK 1588
QY 610 TTVKEFILNKDGE--VSELKPH-----RVTVTIONGKMSSTIIVSEB 651
DB 1589 VAKAKKKLREKVEGVSGLKGVHDEVMKYVQKIDKEVDKEVSKALSKNDVTNVLKQND 1648
QY 652 FILPVYKSELKGYQDFGWEISGFEKGDAGYVNL-----SKDTFIKPVK 698
DB 1649 FFSKV--KNFVKCKVPAAPFISAVAA--FASVYVGFFTSLFSSCVTIASSVYLISKVDK 1705
QY 699 KIEEKEEENK---TPDV-----SKKDNPOVNHQSOLNESHKEDLOR 739
DB 1706 TINKKE---RPPYSFVFDI FKNLKHVLOOMKEKFSKEKNNNVEVT--NKAEKGNVQV 1760

QY 740 EHSQKSDSTKDVTATVLDKNNISSKSTNNPNK 773
DB 1761 TNKTEK-----TKVDKNNKVPKCRRTQSK 1786

Search completed: February 16, 2005, 19:32:38
Job time : 147 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:09:54 / Search time 181 Seconds
(without alignments)

2166.946 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGEIAESKFKNLGNKGEGS.....ATVLDKNNISKSTNNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: uniprot_03:*
2: uniprot_trembl:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	4026	100.0	2140	2 Q97RY6	Q97RY6 streptococc
3	3793.5	94.2	2144	2 Q96AM8	Q96AM8 streptococc
4	3709.5	92.1	2144	2 Q8DQ7	Q8DQ7 streptococc
5	270.5	6.7	2649	2 Q7RAS7	Q7RAS7 plasmodium
6	265.5	6.6	1642	2 Q81B84	Q81B84 plasmodium
7	264	6.6	3504	2 Q81L45	Q81L45 plasmodium
8	260	6.5	1850	2 Q7RGP2	Q7RGP2 plasmodium
9	259	6.4	2757	2 Q7RRR9	Q7RRR9 plasmodium
10	258.5	6.4	1811	2 Q7RRH9	Q7RRH9 plasmodium
11	257.5	6.4	2661	2 Q7RMS4	Q7RMS4 plasmodium
12	257	6.4	1389	2 Q7RRJ4	Q7RRJ4 plasmodium
13	255	6.3	1177	2 Q81BP4	Q81BP4 plasmodium
14	253.5	6.3	1127	2 Q9YVT6	Q9YVT6 melanoplus
15	252.5	6.3	2227	2 Q81I21	Q81I21 plasmodium
16	252.5	6.3	2273	2 Q7RQB6	Q7RQB6 plasmodium
17	251.5	6.2	2849	2 Q81HY4	Q81HY4 plasmodium
18	251.5	6.2	3381	2 Q81E65	Q81E65 plasmodium
19	251.5	6.2	3519	2 Q81E65	Q81E65 plasmodium
20	250	6.2	1474	2 Q81I02	Q81I02 plasmodium
21	250	6.2	1650	2 Q77J28	Q77J28 plasmodium
22	247.5	6.1	5767	2 Q81S28	Q81S28 plasmodium
23	246.5	6.1	2033	2 Q81M18	Q81M18 plasmodium
24	245	6.1	2269	2 Q81UA2	Q81UA2 plasmodium
25	242.5	6.0	2664	2 Q7REL0	Q7REL0 plasmodium
26	242.5	6.0	3317	2 Q8EWP8	Q8EWP8 mycoplasma
27	241.5	6.0	1033	2 Q81BB8	Q81BB8 plasmodium
28	241.5	6.0	2694	2 Q7RJ11	Q7RJ11 plasmodium
29	241.5	6.0	3063	2 Q61MCI	Q61MCI plasmodium
30	240.5	6.0	2740	2 Q7RFS2	Q7RFS2 plasmodium
31	240.5	6.0	4433	2 Q81J15	Q81J15 plasmodium

32	240	6.0	1455	2 Q81KG8	Q81KG8 plasmodium
33	240	6.0	10061	2 Q81J21	Q81J21 plasmodium
34	239.5	5.9	2569	2 Q81BG8	Q81BG8 plasmodium
35	239	5.9	2586	2 Q7PDT7	Q7PDT7 plasmodium
36	238	5.9	1081	2 Q8XIL2	Q8XIL2 clostridium
37	238	5.9	3322	2 Q8XIL0	Q8XIL0 plasmodium
38	238	5.9	3628	2 Q96BY0	Q96BY0 plasmodium
39	238	5.9	3704	2 Q81KY8	Q81KY8 plasmodium
40	237.5	5.9	1104	2 Q7RS08	Q7RS08 plasmodium
41	237.5	5.9	3535	2 Q81C29	Q81C29 plasmodium
42	237	5.9	2511	2 Q81L44	Q81L44 plasmodium
43	237	5.9	5229	2 Q7RTP4	Q7RTP4 plasmodium
44	236.5	5.9	1114	2 Q97242	Q97242 plasmodium
45	236.5	5.9	1738	2 Q81AL5	Q81AL5 plasmodium

ALIGNMENTS

```

RESULT 1
Q9AHT5 PRELIMINARY: PRT: 2119 AA.
AC Q9AHT5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease (Fragment).
GN Name:prta;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
DOI=10.1128/IAI.69.3.1593-1598.2001;
RA Wiseman T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
  Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
  Gayle A., Brewah Y.A., Walsh W., Barron P., Lathigra R., Hansen M.,
  Langermann S., Johnson S., Koenig S.;
RT "Use of a whole genome approach to identify vaccine molecules
  affording protection against Streptococcus pneumoniae infection.";
RL Infect. Immun. 69:1593-1598(2001).
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
  an amide bond (By similarity).
EMBL: AF291699; AAK19159.1; -.
DR HSSP; P00782; ZSER.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; ProC_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFA; TIGR01167; LPXG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
DR Cell wall; Peptidoglycan-anchor; Protease.
KW NON_TER
FT

```

```

SQ SEQUENCE 2119 AA; 238227 MW; 51799B7F6B960A6A CRC64;
Query Match 100.0%; Score 4026; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2e-161;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGELAESFKNLGNGKESGLKDTTGVHHHQQNEBESIKESFTIDRINISTIRDFENK 60
DB 1313 KLGELAESFKNLGNGKESGLKDTTGVHHHQQNEBESIKESFTIDRINISTIRDFENK 1372
QY 61 DLKKLIIKKKFRVDDFTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYTEKLDSEIKSI 120
DB 1373 DLKKLIIKKKFRVDDFTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYTEKLDSEIKSI 1432
QY 121 YGVLSPSKDGHEIILGKISNVSKNAKYVYGNVYSIKATKYDFHSKTMTPDLANIND 180
DB 1433 YGVLSPSKDGHEIILGKISNVSKNAKYVYGNVYSIKATKYDFHSKTMTPDLANIND 1492
QY 1493 IYDGLAFAGDMRLFYKNDNQKKAELIKIRPEKIKETKSEYVYSSYGVNIEIGEGLSKN 240
DB 181 IYDGLAFAGDMRLFYKNDNQKKAELIKIRPEKIKETKSEYVYSSYGVNIEIGEGLSKN 1552
QY 241 KPDNLTMSGKIYDSSEKQOYLKDNIIIRKGVALKVTYVPGKTMLEGNVYSKEDI 300
DB 1553 KPDNLTMSGKIYDSSEKQOYLKDNIIIRKGVALKVTYVPGKTMLEGNVYSKEDI 1612
QY 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLMASLDGNIIRYQVFTFKMDKGBA 360
DB 1613 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLMASLDGNIIRYQVFTFKMDKGBA 1672
QY 361 IDKDGNLVTDSSKLVIFGDDKEVYGEDKFNVEALIKEDSMLEFIDKPPVNLMDKNYFNP 420
DB 1673 IDKDGNLVTDSSKLVIFGDDKEVYGEDKFNVEALIKEDSMLEFIDKPPVNLMDKNYFNP 1732
QY 421 SKSNKIYVNPPEYILRGKISDGKGFNWEALRVNESVVDNLIYGDLIHIDTRDFENIKLVNK 480
DB 1733 SKSNKIYVNPPEYILRGKISDGKGFNWEALRVNESVVDNLIYGDLIHIDTRDFENIKLVNK 1792
QY 481 DGDIDMGMKDYKANGFPDKVTDMGDNVYLQGYSDLNKAVGVHGYFLYDNVKEBVND 540
DB 1793 DGDIDMGMKDYKANGFPDKVTDMGDNVYLQGYSDLNKAVGVHGYFLYDNVKEBVND 1852
QY 541 PKGNISIEYADKSVFENINDKRNNGFDEIOBHITYNGKETSNDIKOIIDKTLNKK 600
DB 1853 PKGNISIEYADKSVFENINDKRNNGFDEIOBHITYNGKETSNDIKOIIDKTLNKK 1912
QY 601 IYVKDFARTTYKEFLINKDTGEVSELKPHRYVTITONGKEMSSITVSEEDFILPYKGE 660
DB 1913 IYVKDFARTTYKEFLINKDTGEVSELKPHRYVTITONGKEMSSITVSEEDFILPYKGE 1972
QY 661 LEKGIQFDGWEISGFEKGDAGVITNLSDOTFIKPVFKKIEEKEEENKPTFDVSKKCN 720
DB 1973 LEKGIQFDGWEISGFEKGDAGVITNLSDOTFIKPVFKKIEEKEEENKPTFDVSKKCN 2032
QY 721 PUVNHSQNLNESHKEDLQREESHQSSTKDYATATLDKNNISSTKSTNNPK 773
DB 2033 PUVNHSQNLNESHKEDLQREESHQSSTKDYATATLDKNNISSTKSTNNPK 2085

RESULT 2
Q97RY6 PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLOCusNames=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=1313;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Knouri H.M., Wolf A.M., Ullrich T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RT Science 293:498-506(2001).
CC -! SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (by similarity).
DB EMBL, AB007373; AAK74791.1; -.
DB PIR, F95074; F95074.
DB HSSP, P00782; 2S8T.
DB MEROPS, S08.064; -.
DB TIGR, SP0641; -.
DB GO, GO:0009986; C:cell surface; IEA.
DB GO, GO:0005618; C:cell wall; IEA.
DB GO, GO:0008233; F:peptidase activity; IEA.
DB GO, GO:0004283; F:subtilase activity; IEA.
DB GO, GO:0006508; F:proteolysis and peptidolysis; IEA.
DB InterPro, IPR010435; DUF1034.
DB InterPro, IPR001899; Gram_pos_anchor.
DB InterPro, IPR003137; PA.
DB InterPro, IPR002029; Pept_S8_S53.
DB InterPro, IPR010259; Pept_inh_S8A.
DB InterPro, IPR001680; WD40.
DB Pfam, PF06280; DUF1034; 1.
DB Pfam, PF00746; Gram_pos_anchor; 1.
DB Pfam, PF02225; PA; 1.
DB Pfam, PF00082; Peptidase_S8; 1.
DB Pfam, PF05922; Subtilisin_N; 1.
DB PRINTS, PR00723; SUBTILISIN.
DB TIGRFAMs, TIGR01167; LPTXG_anchor; 1.
DB PROSITE, PS00847; GRAM_POS_ANCHORING; 1.
DB PROSITE, PS00137; SUBTILASE_HIS; UNKNOWN 1.
DB PROSITE, PS00138; SUBTILASE_SER; UNKNOWN 1.
DB PROSITE, PS00678; WD_REPEATS_1; UNKNOWN 1.
KW Cell wall; Complete proteome; Peptidoglycan-anchor; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; FA44ADB2938B334 CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2e-161;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGELAESFKNLGNGKESGLKDTTGVHHHQQNEBESIKESFTIDRINISTIRDFENK 60
DB 1313 KLGELAESFKNLGNGKESGLKDTTGVHHHQQNEBESIKESFTIDRINISTIRDFENK 1393
QY 61 DLKKLIIKKKFRVDDFTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYTEKLDSEIKSI 120
DB 1373 DLKKLIIKKKFRVDDFTSETGKRMEEYDYKYDDKGNIIAYDGTDLLEYTEKLDSEIKSI 1453
QY 121 YGVLSPSKDGHEIILGKISNVSKNAKYVYGNVYSIKATKYDFHSKTMTPDLANIND 180
DB 1433 YGVLSPSKDGHEIILGKISNVSKNAKYVYGNVYSIKATKYDFHSKTMTPDLANIND 1513
QY 1493 IYDGLAFAGDMRLFYKNDNQKKAELIKIRPEKIKETKSEYVYSSYGVNIEIGEGLSKN 240
DB 181 IYDGLAFAGDMRLFYKNDNQKKAELIKIRPEKIKETKSEYVYSSYGVNIEIGEGLSKN 1573
QY 241 KPDNLTMSGKIYDSSEKQOYLKDNIIIRKGVALKVTYVPGKTMLEGNVYSKEDI 300
DB 1553 KPDNLTMSGKIYDSSEKQOYLKDNIIIRKGVALKVTYVPGKTMLEGNVYSKEDI 1633
QY 301 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLMASLDGNIIRYQVFTFKMDKGBA 360
DB 1613 AKIQKANPMLRALSETTIYADSRNVEDGRSTOSVLMASLDGNIIRYQVFTFKMDKGBA 1693

```

QY 361 IDKGNLVTSSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLMDKNYFNP 420
 DB 1694 IDKGNLVTSSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLMDKNYFNP 1753
 QY 421 SKSNKIYVNPFFYLGRKISDGKGFNMLRVNESVVDNLYIGDLHIDTRDFNKLANK 480
 DB 1754 SKSNKIYVNPFFYLGRKISDGKGFNMLRVNESVVDNLYIGDLHIDTRDFNKLANK 1813
 QY 481 DGDINDMGKDYKANGFPDKVTMDGNNVLTQGYSDLNKAVGVHYQFLYDNVKEPVND 540
 DB 1814 DGDINDMGKDYKANGFPDKVTMDGNNVLTQGYSDLNKAVGVHYQFLYDNVKEPVND 1873
 QY 541 PKGNTSIEYADGKSVFVFNDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 600
 DB 1874 PKGNTSIEYADGKSVFVFNDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 1933
 QY 601 IYVKDPARTTYKEFLINKDTGSEVSELKPHRTVVTIIONGKEMSTIVSEDFILPYKGE 660
 DB 1934 IYVKDPARTTYKEFLINKDTGSEVSELKPHRTVVTIIONGKEMSTIVSEDFILPYKGE 1993
 QY 661 LEKGVOFQDWEISGEGKADAGVYNLSKDTFIKPVFKIIEKKEEENKPTDVSCKKN 720
 DB 1994 LEKGVOFQDWEISGEGKADAGVYNLSKDTFIKPVFKIIEKKEEENKPTDVSCKKN 2053
 QY 721 PQVNSQLNESHKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 773
 DB 2054 PQVNSQLNESHKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 2106

RESULT 3

Q9S4M8 PRELIMINARY; PRT; 2144 AA.

ID Q9S4M8
 AC Q9S4M8
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE Cell wall-associated serine proteinase precursor PrtA.
 GN Name:PrtA;
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3.B.
 RX MEDLINE=21585565; PubMed=11728722;
 RA Zysk G.;
 RA "The cell wall-associated serine protease PrtA: a highly conserved
 RT virulence factor of Streptococcus pneumoniae.";
 RL FEMS Microbiol. Lett. 205:99-104(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (By similarity).
 DR EMBL; AF127143; MAD48399.1; -.
 DR HSSP; P00782; 2S8T.
 DR MEROPS; S08.064; -.
 DR GO; GO:0009986; C:cell surface; IEA.
 DR GO; GO:0005618; C:cell wall; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; F:protease activity and peptidolysis; IEA.
 DR InterPro; IPR010435; DUF1034.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Pept_S8_S53.
 DR InterPro; IPR010259; Prot_inh_S8A.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF06280; DUF1034; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.

DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRFAMe; TIGR01167; LPXTG anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN 1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
 DR Cell wall; Peptidoglycan-anchor; Signal.
 FT SIGNAL 1
 FT CHAIN 20 2144
 FT PrtA.
 FT cell wall-associated serine proteinase

SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match 94.2%; Score 3793.5; DB 2; Length 2144;
 Best Local Similarity 94.7%; Pred. No. 1.3e-151;
 Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;

QY 1 KLGEIAESKFKNLGNGKESGLKQDTTGYEHHHQBESIKESSTIIRNISTIRDFENK 60
 DB 1339 KLGEIPESEKFKNLKVVKQDSLKNETAEVENNLVLDNQSIEGKSLFNHKTISTIRDFENK 1398
 QY 61 DLKXLIKKKFRVDDFTSETGKRMEEVDYKYDDKGNIIAYDGTDLLEYETKLDDEIKSKI 120
 DB 1399 DLKXLIKKKFKQEDPVN-GGRTVYERDVKYDDKGNIIAYDGTDLLEYETKLDDEIKSKI 1457
 QY 121 YGVLSPSKDGHEILIGKISNVSNAKVVYGNVYSIKATKYDPHAKTMTFDLYANIND 180
 DB 1456 YGVLSPSKDGHEILIGKISNVSNAKVVYGNVYSIKATKYDPHAKTMTFDLYANIND 1517
 QY 181 IYDGLAFAGDMFLFYDNDQKAEIKIRPEKIKETKSEYPVYSSYGVNIEIGEGLSKN 240
 DB 1518 IYDGLAFAGDMFLFYDNDQKAEIKIRPEKIKETKSEYPVYSSYGVNIEIGEGLSKN 1577
 QY 241 KEDNLTNRESGKIYSDSEKQOYLKDNIIIRKGYLAKTTYPAGKTDMLBEGNGVSKEDI 300
 DB 1578 KEDNLTNRESGKIYSDSEKQOYLKDNIIIRKGYLAKTTYPAGKTDMLBEGNGVSKEDI 1637
 QY 301 AKIQKAPNLRALSETTIYVDSRVNEDGRTQSVLMSALDGNFIIRYQVTFKMDKGBA 360
 DB 1638 AKIQKAPNLRALSETTIYVDSRVNEDGRTQSVLMSALDGNFIIRYQVTFKMDKGBA 1697
 QY 361 IDKGNLVTSSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLMDKNYFNP 420
 DB 1698 IDKGNLVTSSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLMDKNYFNP 1757
 QY 421 SKSNKIYVNPFFYLGRKISDGKGFNMLRVNESVVDNLYIGDLHIDTRDFNKLANK 480
 DB 1758 SKSNKIYVNPFFYLGRKISDGKGFNMLRVNESVVDNLYIGDLHIDTRDFNKLANK 1817
 QY 481 DGDINDMGKDYKANGFPDKVTMDGNNVLTQGYSDLNKAVGVHYQFLYDNVKEPVND 540
 DB 1818 DGDINDMGKDYKANGFPDKVTMDGNNVLTQGYSDLNKAVGVHYQFLYDNVKEPVND 1877
 QY 541 PKGNTSIEYADGKSVFVFNDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 600
 DB 1878 PKGNTSIEYADGKSVFVFNDRNNNGFDEIOEHIYNGKEYTSFNDIKQIIDKTLNIX 1937
 QY 601 IYVKDPARTTYKEFLINKDTGSEVSELKPHRTVVTIIONGKEMSTIVSEDFILPYKGE 660
 DB 1938 IYVKDPARTTYKEFLINKDTGSEVSELKPHRTVVTIIONGKEMSTIVSEDFILPYKGE 1997
 QY 661 LEKGVOFQDWEISGEGKADAGVYNLSKDTFIKPVFKIIEKKEEENKPTDVSCKKN 720
 DB 1998 LEKGVOFQDWEISGEGKADAGVYNLSKDTFIKPVFKIIEKKEEENKPTDVSCKKN 2057
 QY 721 PQVNSQLNESHKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 773
 DB 2058 PQVNSQLNESHKEDLOREHSHQSDSTKDYATATVLDKNNISSTKSTNNPNK 2110

RESULT 4

Q8DQP7 PRELIMINARY; PRT; 2144 AA.

ID Q8DQP7
 AC Q8DQP7;

DT 01-MAR-2003 (TRENBLREL. 23, Created)
 DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
 DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
 DT Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
 GN Name=prta; Ordered locus names=ap00561;
 OS Streptococcus pneumoniae (strain ATCC BAA-295 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NC NCB1_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RX DOI=10.1128/JB.183.19.5709-5717.2001;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA Dehoff B.S., Estrem S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmore R., Glaes J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McAhren S.M., McHenry M., McMaster K., Mundy C.W., Nicas T.I.,
 RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Roatock P.R. Jr., Skatrud P.L.,
 RA Glaes J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
 RU J. Bacteriol. 183:5709-5717(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 an amide bond (By similarity).
 CC EMBL; AB008434; AAK9365.1; -.
 CC PIR; A97942; A97942.
 DR HSP; P00782; 2SRT.
 DR MEMOPS; S08_064; -.
 DR GO; GO:000986; C:cell surface; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilisin activity; IEA.
 DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR010435; DUF1034.
 DR InterPro; IPR010999; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Pept_S8_S53.
 DR InterPro; IPR010259; Prot_inh_S8A.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF06280; DUF1034; 1.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR Cell wall; Complete proteome; Peptidoglycan-anchor.
 KW KW
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4BDBCS03AOC CRC64;

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
 Best Local Similarity 92.4%; Pred. No. 4, 5e-148;
 Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 KLGIAESKFKKNGKESGLKKDTGVHHQENESIKESSTFDINISITRDPENK 60
 DB 1339 KLGIESKDKAKAKPEENT-NNQAKDBSKPDKSSVSGASJLEINTITITIEFENK 1397

QY 61 DKKLIIKKKREVDPTSETSGKMEBYDYKDKGNIINAYDGTDLLEYTEKLDIEKXI 120
 DB 1398 DKKLIIKKKREVDPTSETSGKRIEYDYKDKGNIINAYDGSALQYTEKFEIKSKI 1457

QY 121 YGVLSPSDGHFEIIGKISNVSKNAKYVGNVYKISLEIATKYDHSKMTMDLYANIND 180
 DB 1458 YGVLSPSDGHFEIIGKISNVSKNAKYVGNVYKISLEIATKYDHSKMTMDLYANIND 1517

QY 181 IYDGLAFADGMLFYVNDQKAKETIKRPEKIKETKSEPYVSSYGVNIEFGEDLSKN 240
 DB 1518 IYDGLAFADGMLFYVNDQKAKETIKRPEKIKETKSEPYVSSYGVNIEFGEDLSKN 1577

QY 241 KPDLITKMSGKIYSDSEKQYLLKDNILIRKGVALKVTTYPGKTMDLEANGVYSKEDI 300
 DB 1578 KPDLITKMSGKIYSDSEKQYLLKDNILIRKGVALKVTTYPGKTMDLEANGVYSKEDI 1637

QY 301 AKIQKANPLALSTTTIYADSRNVEDGRSQSVMSALDGNITIRYOVFFKMDKDEA 360
 DB 1638 AKIQKANPLALSTTTIYADSRNVEDGRSQSVMSALDGNITIRYOVFFKMDKDEA 1697

QY 361 IDKGNLVTDSKVLFGKDKXEYTGEDKFVNEAIKEDGSMFLIDTKEVNLISMDKNYFNP 420
 DB 1698 IDKGNLVTDSKVLFGKDKXEYTGEDKSNVEAIKEDGSMFLIDTKEVNLISMDKNYFNP 1757

QY 421 SKSNKIYVRNPEFYLRGKISDGKGFNWEIYNVESVVDNLYIGDIAHINTRDENIKLVNK 480
 DB 1758 SKSNKIYVRNPEFYLRGKISDGKGFNWEIYNVESVVDNLYIGDIAHINTRDENIKLVNK 1817

QY 481 DGDIDMKMKQYKANGPDKYLTDMGNVYLTQVSDLNAAKVGHVQFLYDNVKEVND 540
 DB 1818 DGDIDMKMKQYKANGPDKYLTDMGNVYLTQVSDLNAAKVGHVQFLYDNVKEVND 1877

QY 541 PKGNTSIEYADGKSVFVFNINDKRNNGFGEIOEQHIYINGKEYTSFNDIKOITDKTLNIX 600
 DB 1878 PKGNTSIEYADGKSVFVFNINDKRNNGFGEIOEQHIYINGKEYTSFNDIKOITDKTLNIX 1937

QY 601 IVVKDFANNTYKEFIINKDTGEVSELKPRVTVTIQNGKENSSTIVSEEDFILPVYGE 660
 DB 1938 IVVKDFANNTYKEFIINKDTGEVSELKPRVTVTIQNGKENSSTIVSEEDFILPVYGE 1997

QY 661 LEKGYPFGWETSGEGKKDAGVYVNLISKDPFIKVPFKIEKKEENKPFVDSKKDN 720
 DB 1998 LEKGYPFGWETSGEGKKDAGVYVNLISKDPFIKVPFKIEKKEENKPFVDSKKDN 2057

QY 721 PQVNHSQLNESHKEDLQREBHSQKSDSTKVATVLDKNNISSKSTNNPK 773
 DB 2058 PQVNHSQLNESHKEDLQREBHSQKSDSTKVATVLDKNNISSKSTNNPK 2110

RESULT 5
 ID Q7RAS7 PRELIMINARY; PRT; 2649 AA.
 AC Q7RAS7;
 DT 01-MAR-2004 (TRENBLREL. 26, Created)
 DT 01-MAR-2004 (TRENBLREL. 26, Last sequence update)
 DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=PY06422;
 OS Plasmodium yoelii yoelii.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCB1_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooil T.W., Pertea M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallow S.J., van Aken S.E., Riedmiller S.B., Feldblyum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoai A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RT Nature 419:512-519(2002).
 CC -1- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL; AABL01002172; EAA18637.1; -.
 DR InterPro; IPR011591; Botulinum.
 DR ProDom; PD001963; Botulinum; 2.

KW Hypothetical protein.
SQ SEQUENCE 2649 AA; 309810 MW; E7207F344643AC24 CRC64;

Query Match 6.7%; Score 270.5; DB 2; Length 2649;
Best Local Similarity 20.6%; Pred. No. 0.0053; Indels 307; Gaps 49;
Matches 206; Conservative 160; Mismatches 327;

1 KLGLAIESKFKNLGN---GKEGSLKDDTGVGHHQHEEESIKEKSFITIDRNIISTID 56
1044 KNNLIERYRRDNNLHQVKA KNNKDI SNLILYNSKSGDNYVQRRSYITETPTETIS 1103
57 FE-----NKDLKLKKKFRVVDFTSETGKMEBYDYKDDKNIAYD----- 102
1104 FPDKIKILQLQNERLKKEIKLEKYDEKEKYNKY-KMEKIKKKQD---NLFEATDGHKK 1159
103 -GTDLEVEREKDELKSKLYGVLSFGDGHFEILGINSVSKAKAYYGN--NYSIEIK 159
1160 LHCELENNKSKONELLKINL-----KDKDMKILEESQICNN-NVSHSEHLNHTLVDVK 1211
160 ATKYDFHSKTMFEDLYANINDIYDGLAFAGDM-----RLFVKDNDQKAE-IKIRMEKI 213
1212 NSENNYKHTNNNDVHSLSDSITDKKKIKKQIVLQDQFLMKLKEIKMEPLSKNVELN 1271
214 KETKSEYVYSSYGAVIELGEGDLSKPKPDNL----- 245
1272 KLLNMKKNHINEYERSIDLEKSIDLKEQLLQLEQICDLKEKNIMEKAAQLRDESS 1331
246 ---FMESG-----KIYSD----- 256
1332 NTTSSISDGTITNNEIKIMEKEIEALYDKIKLKLNSLEBKTNKNITLITLKTISNEOS 1391
257 ---SEKQOYLKDNILIRKGYALKVTYVPGKIDMLGNGVYSKEDIKIQKAPMLRAL 313
1392 IELANKIKLLENKRLQKNEKSI-----NDLKKNIKYD-EDILKENSITKPSIT- 1442
314 SETTIYADSRNVEDGSGTOSVLSALD---GFNIIR--YQVTFEKNKDG-----EAID 362
1443 --TFPLNNEKIEHNTVKQTDDESTQIDDSYDMRIKGYDISNHKONSECVQDQITEID 1500
363 KD-----GNL---VTDSKLYLFGKDDKEYT--GEDKFNVALEKEDSMLFIDTKPV- 409
1501 KMYEKKCTPNLKEITQTDITNLVLSVDKKEMTKKNQDDYN---ILEDNNLKINETNSMF 1557
410 NLSDMKYFNPSKSNKIYVRNPEFY-----LRGKISDPKG---FMWELRVNSVVD 457
1558 NNRKENITNLKNEKKYVANNIFEIDSRANLQMFNSNANGNESFPNAYEVANNE--H 1614
458 NYL-----IYGDLDHIDTRDFNIK--LNVKXGDJMDWGMKDYKANGPDDKVTMDGNY 508
1615 NYTNMKIKINAYSHSLSN--NISEHYSVQNTESLQ-----NVNKKVYKLYAPLEDY 1664
509 YLQGTGSDMAKAVGVH-----YQFLDYNVKPEVNIIDP 541
1665 YNRN-----NLQFISLKHODENNRRTISENGITISYNSVEKDNDSNRYNNSNKEVYIEN 1719
542 KNGTSEIYVADGKSVFENIDK-----RANGPDGEIOEQHIYINGKEYTSFNDIKQIIDK 595
1720 NHRNSIKNSD-----NVNKKYIECLANKINEFESANNESTLANTETENNSTDLKILIE 1773
596 TLNLIK---VVKDPARTTYKEPILN-KDTGEVSELKPRVTVYTIQNGKEMSSTIVS-EE 650
1774 DNNIAVANNKIENY--NDQDLKAYLNSQKTNNNNEHKPND--NMTNEKKKGEAMITIDIK 1831
651 DFLPYRKGELEGYQPDGWEISGFGKQDAGYVNLKDTFIKVPFKIKLEKKEENKP 710
1832 NEVLPTYK--IEKCVLSD-----EKDAGTKKNNK-----KSCQCKNSKNST 1872
711 ---TFDVS-----KKDNPQVH--H-----SOLNEHR 723
1873 VNNMRKRTIYIVRPSBSISIMKTEPANKKAPNLMHKTKNGKLDLSLVNDISKLYANKI 1932
734 KEDLQREHSQKDSSTKDVY---ATVLDKNNISSKSTTN 770

DB 1933 KEELKKNIS-----KQITNLEITTIKKKNPISKSLNN 1966

RESULT 6
Q81B84 PRELIMINARY; PRT; 1642 AA.

AC Q81B84; 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein MAL8P1.29.
GN Name=MAL8P1.29;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844507; CAD51123.1; -
KW Hypothetical protein.
SQ SEQUENCE 1642 AA; 194753 MW; 701F5D28B2BE8097 CRC64;
Query Match 6.6%; Score 265.5; DB 2; Length 1642;
Best Local Similarity 20.4%; Pred. No. 0.0048;
Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;
4 ELAESKFKNLGNGKGSGLKDTGVGHHQHEEESI---KESSFTIDRNIISTIDFENK 60
104 ELNRRNNRRNNRRNNNSNSNFIDKEKEKTNRTIMLKHKRKYPDPTNS-----DEK 158
61 DLKLLIKKKFRVVDFTSETGKRM-----EEDYDYDDKGNIIAYDDGTLEETEKLDE 115
159 NIKR-----KINDIINKYKKGMIIIPSEYDEKYSNGKSGNID-----DE 201
116 IKSKIYGVLSFGDGHFEILGINSVSKAKAYYGNKYISIBIKATKYDFHSKTMFEDLY 175
202 M-----DNKSNVYISGK-----NY-STENNNDKNNFHIRSTL--EKQ 235
176 ANINDIYDGLAFAGDMRLFVKDNDQKAEIKIRMEKIKETSEYFPYSSYGAVIELGSG 235
236 DKINN-----EKEY-----STKOSIDLRRESISHSGKE-----NNNIDLKCH 275
236 DLSKPKPDNLFMESGKIYSDSEKQOYLKDNILIRKGYA----- 275
276 NLIRNKSMLVNPDRNK-----RKENFIPLHNYMKKGNGKNTNINSYDIENIHDEKDK 330
276 ---LKVITYNPGK---TDMLEGN-----GVYSKE-DIAKIQKANPMLRAL 314
331 NEILANNLIDINSYNEBKCLSYNNMKENDFPKIGISKQPNVDSKERKYNKEKOHNIYNDS 390
315 ETTIYADSRNVEDG--STOSVLSALDGFNITRIYQVTFEKNKDG----- 357
391 DDGKSIDRTNYYDNNSSSHVSLSLN--FNESNFMLEDEKIKIDKEVYLIERSLERNS 448
358 -----GEALDKD--GNLYTD-----SSKLYLFGKDDKEYTGEDKFNVAIE 397
449 EYENELKEGETHKDELEKLVLDMTIKDINEKKKVVILYHKQYSE-----NYEKSSIS 503
398 DGSMLFIDTKPVNLSDMKRYFNPSKSNKIYVRNPEFYLRGKISDKGFMWELRVNSVVD 457
504 EN-----IDIGPTNRSGLVDGNNTKQMTIHENNDR-----DKRNDN--INN--D 544
458 NYLIYGDLDHIDTRDFNIKLVNKGQDIDWGMKDYKANGPDDKVTMDGNY-----YL 510
545 NIIINNDDNNIINNANNVNVK--ESNEDIF-----PKANGLSEKINDIVQNIIRIKIKKS 596
511 QGTGSDMAKAVGVHGYQFLYDNY--KPEVNIIDPKGN---TSIEYADGKSVFENIDKRN 565
597 NSGNVNIIDAN-----LSDGISNDEANLKVNTSMDTYKLYYDDK-----GENTON 642
566 GPDGEIOEQHIYINGKEYTSFNDIKQIIDKTLINIKIYVDDFARTTVKFEILNKDGEVS 625

```

Db      643 TF-----KSAAYDKKKLLIENDTKLMHGNNN-----QKEMIRND--EKYLLINGINNSPK 691
Qy      626 ELKPRHVTYIQ-----NGKEMSTIYSEEDFILPVYKGLKGYQFDGEIISGFEGKCD- 680
Db      692 DLNDRNSNDIQVNNNNNNNNNTTFYDAMEYISGNDKEQYNNQVGD- HVIAGVSNKKN 750
Qy      681 -----AGVYINLSKDTFTI-----KVPF-- 697
Db      751 MGTNCSNNNNKNNYINNDRWNHLKEXYFDWMPKPGVDNIIINIKENELVLOKNEKKDIFNK 810
Qy      698 -----KRIEKKKEENK--PTFDVSKKKNPQVNH-----SQNESH--RKEDLQSE 740
Db      811 SHGNEKIEFLKDPNNKNNKVTNIIKKNKVDHIEDVKSVINIGNEIDENKIKKKENILNS 870
Qy      741 EHSOKSDSTKD-VTATVLDKNNISKSTNNEN 772
Db      871 INGEKDIGNKNIINTSYSKNLLSKSYAKEPH 903

```

RESULT 7

```

ID      081L45 PRELIMINARY; PRT; 3504 AA.
AC      081L45:
DT      01-MAR-2003 (Tremblrel. 23, Created)
DT      01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE      Hypothetical protein.
GN      ORFNames=PF14_0404;
OS      Plasmodium falciparum (isolate 3D7).
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=36329;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=2255705; PubMed=12368864; DOI=10.1038/nature01097;
RA      Gardner M.J., Hall N., Fung E., White O., Berthman M., Hyman R.W.,
RA      Carlton J.M., Patil A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA      Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Xyes S.,
RA      Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA      Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA      Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA      McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA      Venter J.C., Carucci D.U., Hoffman S.L., Newbold C., Davis R.W.,
RA      Fraser C.M., Barrett B.,
RT      "Genome sequence of the human malaria parasite Plasmodium
RT      falciparum."
RL      Nature 419:498-511 (2002).
DR      EMBL; AE014822; AAN37017.1; -.
DR      HSRP; Q93IC2; LMKX.
KW      Hypothetical protein.
SQ      SEQUENCE 3504 AA; 408308 MW; B8454D8D55BE4F0 CRC64;

```

Query Match 6.6%; Score 264; DB 2; Length 3504;

Best Local Similarity 21.2%; Pred. No. 0.014; Mismatches 346; Indels 252; Gaps 43;

```

Qy      11 KNLGNGKSGSLKCKDTGVGHHQEN--ESISKEK-----SSFTIDRNISTIDPENKDL 62
Db      1432 KNSHNSSENMITEEYKGNKSPQENIDNIIPEKNGENKNSQNSQNDHNMTEKNEKNS 1491
Qy      63 KKLKKKREVDVDFSETGKMEYDYKDK---GNIIAYDGTDLLETETKLDKDEIKS 118
Db      1492 QQ-----NSDHNIMTEEYKKNNSQENTDNDIMTEGKKNKSEKKEEDIAS 1539
Qy      119 KIYGV-----LSPSKD-GHEILIGKISNYSKNAKYVG-----NNKSIETIKTKDF--- 165
Db      1540 --YELDKNRISHENQEHFTYPSRINKEFTNIVYNIVSGNEKGLSVKNISDIIIPD 1597
Qy      166 -HSKMTPTDLYANINDIVDGLAFAGDMRLFVNDQKAEIKIRMPKIKETKSE---YP 221
Db      1598 GKGKNIQDIIIEEN-----GKNKFEENIEDKISDKTQ--KSKISHNAGCHTF 1647
Qy      222 YVS-----SYGNVIELGEGDLSKNRPDNLTKMESKITV-----SDSEKQVYLL 264

```

```

Db      1648 YESGNKISDENDEVYNISDINTNDOE--EYESKRIFETNDINKHISSDNNKINOK 1705
Qy      265 KONILIRKGYALKVTVYNGPKTD-MLEGVYSEKDKIAOKAPNLR-----ALSETTI 318
Db      1706 QNNIINESQDKNIIVNH--KLDKILKHEHTSDESLKIKENKNTKKNKSINNEKI 1763
Qy      319 YADSRNV-----EDGRSTQGLMSALDFNIIRYQVTFPKN----- 355
Db      1764 EEEKENVKNDFTIIGKKEENESDCLKIOKISNENIKNIIIVTDVYNDKSYNAOGTGH 1823
Qy      356 -----DKGEALD-----KDNLY----- 368
Db      1824 ENDETNGTINISNDGLDNVAKIDQYISKEENILQENKEDIIIPSYTINISLGDYKENS 1883
Qy      369 -TDSKLVLFQKDDKEYTGEDKFNV-----EAIKEDSGMLFIDTK-----PVNLS- 412
Db      1884 PEDIKMEVAHKNIQNTISEDELGTQKDNREKEDKSPNGVEBHNGENDKIGEVNLSN 1943
Qy      413 MDKNYFNPSKSNKIYVRNPEFYLRGK-ISDKGFNWELE--VNESVDNY-----L 460
Db      1944 MNMNESTIGNSDTI--NQHLLNEGKNIHHKGNVNSETNEMTNGTQNIISNEQFEKNI 2000
Qy      461 IYGD.IHINTDPRNFKLVKNGDIDMNGMKDYKANGFPDKYTDMDGNVYLLQGYSDIAK 520
Db      2001 IRGD-DIKDKMKNENYKIEDETGN-----NIKINKYDNKAKILNELIIKQGTDDSDAD 2052
Qy      521 AVGVHYQFLYDVK--PEVNIDPKGNTSLEYDGKSVFNINDKRNNGFGEIOEHIYIN 579
Db      2053 DISTGSDKMDQIEKNENIKHKNVNV-----KDKI--SNDKENIYVPE-HKHDHIS 2105
Qy      580 GKEYTSPNDIKQIIDKTINIKIYVDPAFNTVKEFIIINLKDTGEVSELPKPRVTVIING 639
Db      2106 DNKKKEFNVNLEIPKGNH-----LDCKETITQVEEKSIGQDKSMENNVSF--NDG 2157
Qy      640 KEMSTIYSEDFILPVYKGLKGYQFDGEIISGFEGKCAAGYINLSK----- 689
Db      2158 KDIIH---IOEBDI-----KGNINNVNDKHSKSKNNLHIDBNKYVEKEIKH 2204
Qy      690 ---DTFIPVPFKIEEKKKEENKPT-----PVSKKKNPQVNHQSLNESHKEDLQREE 741
Db      2205 EIAHDIDKEKEIOENDSNKPESENENIIVDNAQDDK-----NISKLTDIHQEE 2256
Qy      742 HSQKSDSTKDYATVLDKNNISKS 766
Db      2257 KGTNNDSDVE-----HNVS DKT 2273

```

RESULT 8

```

ID      07RGP2 PRELIMINARY; PRT; 1850 AA.
AC      07RGP2:
DT      01-MAR-2004 (Tremblrel. 26, Created)
DT      01-MAR-2004 (Tremblrel. 26, Last sequence update)
DE      Hypothetical protein.
GN      Name=PY04304;
OS      Plasmodium yoelii yoelii.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=73239;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=17XNL;
RX      PubMed=12368865; DOI=10.1038/nature01097;
RA      Carlton J.M., Angiuoli S.V., Suh B.B., Kozi J.T.W., Pereira M.,
RA      Silva J.C., Ermolaeva W.D., Allen J.B., Selengut J.D., Koo H.L.,
RA      Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA      Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA      Cho J.K., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M.,
RA      Florens L., Yates F.R. III, Raine J.D., Saindi R.E., Harris M.A.,
RA      Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA      van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA      Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,

```



```

QY 186 AFAQDMRLFVKNDQ-----KKAELIKIRMPKI-----KETSEYEVSSYGVNIEL 232
D 773 -----YAKENDQANVYKSKISIEIRHYNQDISIDNTKEBEAKQNYDSKETYTTISI 824
QY 233 GEGDLSKN-----KPDNLTKME-----SGKIYDSBKQOYLKNDIILRKQALAV 278
D 825 KEDETSKITNEVKNNMDEFLSKYDKYKINEENCKENVDEHNTQFELTKIKAEVSEDTL 884
QY 279 TTY-----NPKTMDLENGVYSKE--DIAKIOKANPNLRALSET-----TYAD 321
D 885 SIYEKENDSKSLINETNNSIEREYGNITLKVDEYIVCSSTESIKGNHNNKITLKE 944
QY 322 --SRVEDGRSTQSVLMSALDGFNIIRYOVFTFKNDKGEALDKGNL---VTDSKVLV 376
D 945 KLNQNIKITIKETNSIEKSYDKFE-----NTLJINKIETLDPKAF--KQASLNDYSSNNELMQ 999
QY 377 PEKDKVEYGEDEKFNVEALKEGDSMLFID-----TKRVNLSMK--NFFNSKN--KIYVN 430
D 1000 YFNNLKNANJKNKEN-----MLYHOFDEKEKAVNNIIOKIEDINKINENIELAIYT 1050
QY 431 PEFLYNGKISDKGFNMEL-----RVNESVVDNLYIGDLHIDTRDPNLIKLVKQDG 483
D 1051 SIYNSIEELIENELGKNIESTLNTNVEKVTNTNPFKIKELKLYNFSDFG-----KEGN 1105
QY 484 IMDWCKDYKANGFPDKVTMDGNVYLTQYSDLNAAKAVGVHYQFLYDNVKEVNI DPK- 542
D 1106 I-----KYTNEVVKIKODIKVNOQIDHHINELIEKKSSENVYDEMKAOINLEKY 1157
QY 543 -----GNTSIEYADG--SVFNNINDKN-----NCFDEIOEQHIYNGKETSFNDK 590
D 1158 ADTAISNDVNEGIEKKQONI VTKIDKKNIYEINLSEISKIE---KOQTSIEKY 1212
QY 591 QI-----IDKTLN 598
D 1213 GINLSYGVNLGTLFLEQIDBEKKKSEHTIKAMEAVMEDLNKKKSQEIENEMGEMDIN 1272
QY 599 IKIIVKDFARNTTVKPEFLNKDTGE--VSELPKPRVTVTIONGKMSSTIVSEDFILPVY 657
D 1273 KEMEVLTNISHDDKCYHTSKHNENISDIRKSLTKI--IQD-----FSRESIDINDI- 1322
QY 658 KELEKGY---QFDGMEISGFEGKXQDAGVINLSKDTFIKPVFKKIEKKEENKPTPV 714
D 1323 KEELQNVSESQKHSEINQVLSK--IANIYILKLKIKIKINKKYEYNTIEK----- 1375
QY 715 SKKKNPQVNHSQLNESHREKEDLOREHESQKSDST---KDVATATVLD---KNNISKSST 767
D 1376 NKNINDELISEKLIKIKIKIDISLKECKSKIESITLDDKDIDGCIQINIKVLTHTISEET 1435
QY 768 TNN 770
D 1436 NIN 1438

```

```

RESULT 10
Q7REH9 PRELIMINARY; PRT; 1811 AA.
AC Q7REH9;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein (Fragment).
GN Name=PY05085;
OS Plasmodium yoelii yoelii.
OC Karyocota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguilo J.S.V., Suh B.B., Kooij T.W., Pearce M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selegue J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldlyum T.V.,
RA Cho J.K., Quackebush J., Sedegah M., Shoabdi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser L.W., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.,
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RT Nature 419:512-519 (2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBD whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001591; EAA17051.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 1811 AA; 210756 MW; E76A90F7392BCA0A CRC64;

Query Match 6.4%; Score 258.5; DB 2; Length 1811;
Best Local Similarity 21.7%; Pred. No. 0.011;
Matches 179; Conservative 117; Mismatches 312; Indels 215; Gaps 40;

QY 11 KVLGKGGSLKQDTTGVHHQENES-----IKKSSFTIDRNISTRDFENKD 61
D 555 KNINNSQHSNIEVNNRNIHIDNNNDKNTVSNNIHLKNDETKIDINVDNANTYTGFE 614
QY 62 LKKLKKKREYVDFTSETGKMEBYDYKDDKGNIIAYDDGTLEY-----ETEKLE- 115
D 615 ILS-DKEDPKLEFLNDKMGSKFLPEKMLDCSKWQIMDLQREYINNNIELKKEE 673
QY 116 -----IKSKIYVLSBQDGHFEIIGKISNVSKN-AKYVGNVYSIEIKATKYDPH 166
D 674 KKKYLEKICINNKIKELMCGRYVGLNV--KLCNMKKKIANEFSLNIIKNHVEVTLTYE 731
QY 167 SKTMTF-DLYANINIDVGLAFAQDMRLFVNDONDKKAEIKIRMPKIKETSEYVYSS 225
D 732 RSAMCYQDLIA-----QLYVIGLANK--KLEMKIKINEK----- 764
QY 226 YGNVIELEBGLSKNKPDLTKMESGK-----IYDSBKQOYLKNDIILRKQAL- 276
D 765 --NNID-GVANGQENESDAKKKQDSKKKGYVHSLIYGFNKKSH--KQN--KKYGSIN 815
QY 277 ---KYTTNPGKTMLBGNVYSKEDIKQKANNLRALSTTTIYADSRVVEDGRSTQS 333
D 816 NDTKQTT--EPSKCD-----NLKKKENERITANNNEDEKIKEN 853
QY 334 VLMSALDGFNIIRYOVFTFK--WNDK-----GEAIDKGNLVTDSKVLFGKD--KEYT 385
D 854 KQKG-----NNTENTVNTKKNLNEQYKKRNSDEDTKRSNIRTDKANNRPFGRNSYNET 908
QY 386 GEDEKFNVEAIKEDGMLFIDTPVNLSDMKVFNPS-KSNKIYVNPPEF---YLRGKIS 440
D 909 VEDK-----EIKETKNSVNEYYSNNSYENNSYLDNNSFKKNLKNKSEIV 953
QY 441 DKGGFNMELRVNESVVDN--YLIYGDLDHIDTRDPNLIKLVKDGIMQMDYKANGP 498
D 954 EKSNNSN---IKKKYVDGGGSGFFGSF-----FMFK-----SDYKKGK-- 987
QY 499 DAVTMDGNVYLTQYSDLNAAKAVGVHYQFLDVKPPEVNI DPKGNTSIEYADGKSVFN 558
D 988 SQISNNSAIVPEPSDSNSN---GSDYSFKYNNKEDKQKQKNSKSNNSAFAFK-IFN 1043
QY 559 INDKRNNPDEIQ---EQHIYNGKETSFPNDIKQIIDKTLNIRIKIYVDFARNTTVKEF 615
D 1044 LSRKKKKKKQGGKPSLSKSEDKIKKKDKNTDNGSVNENDDK-----YNNESSNQ 1091
QY 616 IANKDTGEVSELKPRVTVTIONG-----KEMSTIYSEDFILPVYKGELEKYQFPG 669
D 1092 SSNEEDGHINNYKKN-----NGNKKRANKNGSGNINLNNKSSSLAEERFDEEFQOI-- 1142
QY 670 WEISFEGKQAGVYINLSKDTFIKPVFKKIEKKEENKPTFVDSKKKDNQVNHSQLN 729
D 1143 -----LKREYGDHILTEDVKSITIFERFDEDE-----FDIGTKNSSSCSNS--N 1187

```

QY 730 ESHRKKEDLOREHESOKSDSTKVAT-VLDKNNISSKSTNNP 771
 DB 1188 DSNDSD-----SNKSIROSGDTGTQITKSSSSSKSEKKP 1224

RESULT 11

Q7RMS4 PRELIMINARY; PRT; 2661 AA.
 ID 07RMS4;
 AC 07RMS4;
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Rhoctry protein (Fragment).
 GN Name=PY02104;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxId=73239;
 OK NCBI_TaxId=73239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angluoli S.V., Suh B.B., Kooij T.W., Petrea M.,
 Silva J.C., Ermolaeva M.D., Allen J.E., Sengut J.D., Koo H.L.,
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
 Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,
 Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 Cunningham D.A., Preiser P.R., Bergman L.W., Valdiva A.B.,
 van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 Carucci D.J.;
 RA "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RT Nature 419:512-519(2002).
 RL Nature 419:512-519(2002).
 CC -!- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL/AA01000575; EAA21526.1; -!
 DR EMBL; IPR011561; Pox_11_rel.
 DR InterPro; IPR006499; ReticulocyteBP.
 DR ProDom; PD014111; Pox_11_rel; 1.
 DR TIGRFAMs; TIGR01612; 235Rda-fam; 1.
 FT NON TER 2661 2661
 SQ SEQUENCE 2661 AA; 312269 MW; 0BD02AD0B712253B CRC64;

Query Match 6.4%; Score 257.5; DB 2; Length 2661;
 Best Local Similarity 20.0%; Pred. No. 0.018;
 Matches 179; Conservative 151; Mismatches 294; Indels 271; Gaps 41;

QY 22 KKDTGVHHQNEESIK-EKSSFTIDNIS-----TRDPENK-- 60
 DB 427 KKKLETTITLANKNEETVLEKEIRLFKKYSDEVAEKGVBELKAKETIKDYNKKE 486
 QY 61 -----DLKKLKKKREVDFTSETGKMEBYDYKDD-----KGNIAVD 101
 DB 487 YIKKALDKKALENNKKYIDELGKNTPPQIIEYVKKDITITSTIKSELSEIKGNIVELY 546
 QY 102 DGTDLFEYFEKDEYSK-----ITGVLSPPSGHFEILGKISNYSKN----- 144
 DB 547 NELSSVVGQNTIDPIPKKELETLKSEINDVYNNKIQNMKIEVEELTKNIEYSKELSNL 606
 QY 145 -----AKVYVG--NNYSIEIKATKVDGHSKWTGFDLYANINDVDSLAAGMRPLPVKN 198
 DB 607 ILEIITITGEIDN-----DLNKLKEFKNKEQ--ELSKIND-----YTKEN 647
 QY 199 DQ-----KKAIEKIRMPERI-----KETSSEYVYSSGVNLEIGEDLSK----- 239
 DB 648 DQLSVYQSITIEIRKHYNDQINVDNTKGAEAKONTQOSKEVAKKIPITKGEYSKFINETV 707
 QY 240 -NKPDLTGMF-----SGKIYDSKQOYLKDNIT-----LRKGYALKVTTYNG 284
 DB 708 AMKNELRLTKKYIDPDGAYKKEKANSHEKITELVKKIRIEISKYLVNYYEKK--FNNS 764

QY 285 KTDMLGNGVYSKE--DIAIKOKANPNLRALSETTIVADSRVVEDRGSTQVLAALD-G 341
 DB 765 KLLIITETDSIEBEKKNINTLAKVDYIKVCYHT-----NELISFRNKQTITLKKLDKN 819
 QY 342 FNIIRY-----QVTFK-----MNDGEALDKDGNLV-----TDSSKLVLF-----G 378
 DB 820 INTVKENNSIDGLYNNKENILTDKTELDKFTDVALNDHRSNNNELMKYFNNLKANLG 879
 QY 379 KDDK-----EYTGEDFNVEATKEDGSMLEFIDPKPVLNSMDKNYFSPSKNKIYVNPERY 434
 DB 880 KKKKMLYQGFGEKKALIDDIKKKADINENYSKIEVALIYASINISDMEEREIEKSIES 939
 QY 435 LKGISDKGFMELRVNESVVDNYLIYGDLIHDTPTFNILQANQDGIIMQMKMDYKA 494
 DB 940 INTQVIER-----VKTNTVNLNEIKGLKQYFHPGKEENK----- 977
 QY 495 NGFPDKVTDMGNAVYLGQYSDLNKAVGVHYQFLYDNVKEPVNIDPKGNTSIEYADGS 554
 DB 978 --YPDINKIKNEI-----KTVG---QIDQHNKLEIKKNSGSXNDKIGKT 1021
 QY 555 VVF-NINDKRNNGFGEIQEHIYNGKEYTSFN-DIQIIDKTINIKIVVDPARNTV 612
 DB 1022 DKFENVTDK-----TLYNKDPKEIEEKIKNI-VTKIDKKNYI 1059
 QY 613 KEFLINKDTGEVSELKPHRVT-----VTIQGKEMSSTIVSEDPFILPVYKGELEKGY 666
 DB 1060 K--INKLDEISEIEKNNTSLAKKVDINVSQGLGLFLQID----- 1101
 QY 667 FDGWEISGEGKQAGVYINLSKDTFIKPV-FKKIEKKEEENKPTD---VSKKDN 720
 DB 1102 -----EEKKGAGHMK-SMEYMDLDNKKKSQIEENMKINMDIKYINKEMA 1151
 QY 721 PGVNSQLESIRKEDLOREHESOKSDSTKVATVLDKN---ISSSTNNP 772
 DB 1152 INISHDDKQYH--NISKGEKISD-----IHKNSLIKITQEFSTESNIN 1194

RESULT 12

Q7RPS4 PRELIMINARY; PRT; 1389 AA.
 ID 07RPS4;
 AC 07RPS4;
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=PY01465;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxId=73239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angluoli S.V., Suh B.B., Kooij T.W., Petrea M.,
 Silva J.C., Ermolaeva M.D., Allen J.E., Sengut J.D., Koo H.L.,
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
 Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,
 Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 Cunningham D.A., Preiser P.R., Bergman L.W., Valdiva A.B.,
 van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 Carucci D.J.;
 RA "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RT Nature 419:512-519(2002).
 RL Nature 419:512-519(2002).
 CC -!- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 CC EMBL/AA01000391; EAA20802.1; -!
 DR EMBL; IPR011591; Botulinum.
 DR ProDom; PD001963; Botulinum; 1.

KM Hypothetical protein.
 SEQ SEQUENCE 1389 AA; 164816 MW; 21851C4A630F5D9F CRC64;
 Query Match 6.4%; Score 257; DB 2; Length 1389;
 Best Local Similarity 20.4%; Pred. No. 0.0089;
 Matches 193; Conservative 139; Mismatches 288; Indels 326; Gaps 48;

17 KEGSLKDTGVVHHHOF-----NEESIKESGFTIRNISTIRD 56
 354 KKEAQQEKESSEHLEKNI FYEKEILEASKYDVSYFKNKEIKKDKSNNYNNKGVLP 413
 57 FENKDLKKLKKKFRVDDFTSETGRMEBYDYKDDKNNIAYDGDIDLEYETKLD 116
 414 EINKKQKQKINKKCNKVFESHSDT---EKTEY-----SNLIS-----DVLNQKDK 457
 117 KSKIVGLSPSKDGHFEILKISNYSKNAKVVYGN-NYSIKRIKAT-----KY 163
 458 KNTIYGL---SKEDY---SNCKKQDTEVEYVYDNDKTDIDIQVNDKAVLDISPETHY 510
 164 DHSK-TMTFDLYANINDIVDGLAFAGDMRLVVKNDQKKAIR--MPEKIKETSE 219
 511 KVRSNIIITINKIKNIKIEKI-----EINLITVQLKIPQRIENVLIPK----- 556
 220 YVYVSSYGVN-----ELGEGDLKSKKPDNLTKMESKITYSDEK-----QQYL 263
 557 ---YSMDILKIQIKTISENENDIKTNKE---TKNATSYQOOSDKNSTQPSDNKV 607
 264 LKDNIIILKRGVALKV-----TTYPGKIDMLEGNGVYSKED- 299
 608 IKKMIDRTGSIKCYVYLPLNNINHEILNRILINSTYR-----MYEKML 657
 300 ---IAKIOKANPNLALSETTIYADSRNVEDGRSTQSL-MSALDGFN---IIRYQ-V 349
 658 LPDITLYLQ--SKSISSEIKETMYISIKCVANGIYLYPLINRNVINDPKNPMINSQNN 715
 350 FTFK-----NMDCGEAIDKD-GNLTWDSKLVLFGKDKKEYTGEDKFNVAIKEDGSL 402
 716 STFKPPLVSLYNDL---IDREIGNLI-MTKLI-----NKIKIKKLPKRGYI 759
 403 ---FIDTPKPVNLMSDKN-----YFNPSKSNKIY 427
 760 INDICGRPFDDKNNVLVCVDNRKNSGINQIGQNLIDHIGHETANYIFILMSNFI 819
 428 VAMP-----EFTYRGKISDKGKGNMELRVNESVYNDYILGYDLHDN 469
 820 VKSIDLFPESHLPILBOVYVDFPIEKMLKLPIMFLAITETACIMLCFL----- 872
 470 TDFPNIKLNVKGDIMDMGKDYKANGPPDKVLTDMGNYVLTGYSDLAKAVGYOFL 529
 873 ---FINNAQI---DSLFT---KNLKAVALKDENDQNNFFTLTYTSIE-RVQNNKHCYF 923
 530 YDNVPR-----EVNIDPKGNTSIEYAD--GKSVENINDKRNNGFGEIOQHLYI 578
 924 LNVSPFLIPRFQSEMLN-----NISHADVEGHNLNGTSDERIKKEEIKKKKE 976
 579 NKEKTSFNDIKQIIDKTLINIKIVKDPKRNNTVKEFIANKDTGEVSEKLPKRVVTION 638
 977 NG-----IK-LKKTINDNDFIYDVARQSI-FLYN-----AIKNDALAKNIEN 1018
 639 GKEMSTIVSEDF-----ILPVYKGELEK-----GYQF-----DCM 670
 1019 NYDKKKQIYYTTSYTKQFQDYINILNRYNOLLQKEKKEYSFRREVETIQSINRDOQ 1078
 671 EISGEGKKDAG-YVINLSKOTFIKPVFKKIEKKEE---ENKPTPVSKKKQNPQV 724
 1079 DVKSYFGKNDNNIIFSGGLNQ-----FDKCKNEKREBGIKSSNNKEMQDNKYTYNQNS 1131
 725 HQQLNESHKEDLQREHSHQSDSTQDVATVLDKNNISKSTTN 770
 1132 KQTFSDKKEQNIYTDENIQLN-----NPFNTNQVANN 1165

Q813P4
 ID Q813P4 PRELIMINARY; PRT; 1777 AA.
 AC Q813P4;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein PFEI095w.
 GN Name=PFEI095w;
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxId=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
 RA Buncue C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corron C.,
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphray S., Jagsels K., James K.D., Johnson D., Kerhornou A.,
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitz E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G;
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
 RL Nature 419:527-531(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AL929353; CAD51583.1; -
 DR InterPro; IPR011591; Botulinum; 1.
 DR ProDom; PD001963; Botulinum; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 1777 AA; 213320 MW; 244467CFE190C522 CRC64;

Query Match 6.3%; Score 255; DB 2; Length 1777;
 Best Local Similarity 20.7%; Pred. No. 0.015;
 Matches 186; Conservative 152; Mismatches 324; Indels 226; Gaps 44;

4 EIAESKFNPLANGKSGSLKDTGVVHHHOFNE-----ESIKESGFTIRNISTIRD 57
 977 ELLDDKKKL--DEENELDDKK--KKLDEENELDDKKKLDEENELDDKKKL--DE 1030
 58 ENK--DLKLI-----KKKFRVDDFTSETGRMEBYDYKDDKNNIAYDGT 104
 1031 ENELDDKKKLDEENELDDKKKLDEENELDDKKKLDEENELDDKKKLDEENELDDKKKL 1090
 105 DLEYETKLDLKSXI---YGLSPS-KDGF-----EILGKISNYSKNAK----- 146
 1091 FLOTENHLNNEKTKIQDYDNIIEALEKEMSKKILAEETKIKIENDENDIKRQSQIEN 1150
 147 VYGNYSKIEIKATKYDFHSGKTMFTFDLYANINDIVDGLAFAGDMRLVVKNDQKKAIR 206
 1151 IYRRNSMMINDINISYSKNTKTFNPDNIEE----- 1184
 207 IRPEKIKETSEYPPYSSYGVNIELGSDLSKKNPDNLTKMESKITYSDEKQOYLKD 266
 1185 ---EDKITQKNTKIYTDISNM-----LTKNKSSIYINSSEKDIINEQEIYSKN 1234
 267 NI-----IIRKRYALKVTTYNPGKIDMLEGNGVYSK-----EDIAKIQ-AAPN 309
 1235 NTFNNDIEKDLTNWEKKNNNFNNTSYKKQSNIVDINDSNKFLMDTESIIKIQNDINEK 1294
 310 LRALS--FTIYADSRNVEDGRSTQSLM-SALDGFNIIIRYQVFTKNDGGEALDKDGN 366
 1295 VKLLKQREIYEEKCNFEKKNELHLKENVLANKNIIKOR--ENELNKEEELKQKKE 1352

```

QY 367 LVTDSKLVLF-----GKDDKEY-----TGEDKFNVAIKEDGSMLEFIDTKPVNL5M 413
DB 1353 ILLSKENIINYSLSNKKINDKLEYNKLEBVTKKGDIINMSTMSF5SHL-SDTN--NIQ 1409
QY 414 DKNYFSPSSKNTKYVNPPEFYLNCKISDKGCFNWEELRV-----NESVDNLYLY-G 463
DB 1410 FKNDIN---NDKLYENISSY--GNMLSHDNMKYDNNLSGDNMNSHDNNMARDNICH5N 1465
QY 464 DLHIDTRDGNILKLVN-----KDGLMDGMKDYKANGPPDKVTMDGNVY----- 509
DB 1466 NLSYDNPSTNTMKTJSELENNENKKNIDY---YDKDNFVDNKNIIINHNPNISKI 1521
QY 510 ---LQTSYDLNKAQVAVHYOFLYDNYKPEVNIDPKGNTSIEYADGKSVFVINDKRNNG 566
DB 1522 SNDSISGMDNMB-----LHDNL-----NNSLKEIEMVSKLSKRSDEINT 1562
QY 567 FGEIOEQ-----HIYINGK---BY-TSPNDIKQIIDXTLN 598
DB 1563 LKLIKIKQTEEEKKONNINQINDISSDMNIIYINGNISGVNYDTSMHVLPEPKRFLN 1622
QY 599 IKIVNDFPANTVKEPFIKNDG---EVSLEKPHAVTYTIOGKEMSSITVSEBDFI-- 653
DB 1623 LS-EVED---NEKMKISLSNELVSLYKEINNIK-----EENINVLKKNFIGN 1667
QY 654 -LPVYKGELEKGYQDFGMEISGEGKKAGYVNL5KDTPIKPVFKIEKKEENKPTF 712
DB 1668 LLLNFINDLRNRYL---KENYKEVHKSQIISEREDIKELQMTLNEKJKE----- 1719
QY 713 DVSKKDNPOVNH5QNLNESHK---KEDLQREH5QKSDSTKDTATVLDKNNISKST 767
DB 1720 -ISYKMKLLKN--QINDTYKLKGRKSLSTVELLKQ--DIKFLNEBVLKKEKAVTLST 1772

RESULT 14
Q9YV76 PRELIMINARY; PRT; 1127 AA.
ID 09YV76
AC Q9YV76;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein MSV156.
GN Name=MSV156;
OS Melanoplus sanguinipes entomopoxvirus (MSPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99102612; PubMed=9847359;
RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
DR EMBL; AF063866; AAC97677.1; -.
DR J. VIROL. 73:533-552(1999).
DR PIR; T28317; T28317.
KW Hypothetical protein.
SQ SEQUENCE 1127 AA; 134265 MW; F185DA1D5A3FE7D CRC64;

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0098;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;

QY 7 ESKFKVLGNGKESLKKDITGVHHNHOENESIK--EKS6FTTD--KNISTIRPFKKDK 63
DB 196 EIEFKKIDVQKINKODELKLDBSKKEFKKQEBELKTIQKQEBELIKGLNDKEIN 255
QY 64 KLIKKKFREVDPTSETGKRM-----EYDYKDDKGNIIAYDGT--DLEVERE 111
DB 256 FNIDKQKLLDQNSKINTLNENIKGVMLYTETKOKISLQWELINKOSTISLDEKQ 315
QY 112 KADEISKIKYGVISPEKQGHFELGKISVSK-----NAKYVGNVYSITIKAT 161
DB 316 LDELDKNINNTISLVNKS--TKITVIOQLLESLLDPNNANI-----NINELSKIK 367

```

```

QY 162 KYDFHSKTVTPFLYANINDIVDLAFAGDMRLFVK--DNDOKKA--EIKIRMEKIKET-- 216
DB 368 LPDNDIQKLANDITBQNNKITD--FPNNSTRIFKEGLDIEYKKIDITKNNILQKLESYK 425
QY 217 ---KSEY---PVSSYGNVIELGEGDLSKKNPDNLTKNSEKTIYSDSKQOYLKDNII 269
DB 426 KIDQETEVYKKNIKNEYNDIIEIKNNNLQKLEENKIKIDQETEVYKKNIKNEY--NDIIE 483
QY 270 LKQVLYLTKTYTNP-----KTDMLEGNVYSKEDIKAKQKPNLRALSETTYAD 321
DB 484 LKNNILQKLEENKKNINDKLYLKNDIESNTELFNGLNISDRKDSREIRAKLN--TEYBQ 541
QY 322 SRN--VEDRSTQSVL-----MSALDGNNIIRYQVFTFRMDKGAIDKQGLVYDSK 373
DB 542 LKQDLLENINKNTLMLKLSDKLSLE-----QLYDSKKNIL-DGIDIKYNSLAKEN- 592
QY 374 LVLFQKDKKEYTGEDKFNVAIKED--GSMLEFIDTKPVNL5MDKNYFN--PSKSNKIY 427
DB 593 ---DKIDYFNSIEKFIDYVNIENKFNIGLDSIINKIINNDOFKKEYINSKIDSKEELS 648
QY 428 VNPPEFY-----LRKISDKGCFNWEELRVNESVDNLYLYGDLHIDNTD 472
DB 649 TWFDDIFNAKQIASITNNIENISNKKID--LNEPILSNEDSSKEL-LDEIRK 698
QY 473 FNIKLVNKGDDIMDMGMDKYKANGPPDKVTMDGNV-YLQTSYDLNKA-----V 522
DB 699 YKQQPD-KIKDMANTEVSFE--NTLOKQIDSKSNINELTNAYDIINTAANDLDDGLANN 756
QY 523 GVHYOFLYDNYKPEVNIDPKGNTSIEYAD--GKSVFVINDKRNNGPGEIOEQHI 576
DB 757 GSEFKLVYNNASDLDLDTIQKNDDEKVKQNEYLEKKNQISIEINDIVNN----- 805
QY 577 YINGKRYTSFNDIKQIIDTILNIKIVKDPARTTYKEFIANKDGEVBE--LKPHRYT 633
DB 806 FL--KELIKFNNTT--TNKSLN-ELLTND--DINDIKPKLYKELKIKSTNNILKTIYK-- 855
QY 634 VTIONGKEMSSITVSEBDFILPVYKGELEKGYQDFGMEISGEGKKAGYVNL5KDTPI 692
DB 856 NEIDVNEKLSIVIELOPINSFLEPNQ-----SITSH-----INFLANTLA 900
QY 693 -IKPVFKKIEKKEENKPTFVSKKKNPOVNH5QNLNESHKEDLQREH5QKSDSTXD 751
DB 901 GINDVLYNKLNLKIMADTTRRGDTNR-----DEIKNQISSENIK5QKPFKEKN--KD 950
QY 752 VTATV-----LQKNNISKSTTN 770
DB 951 LKKLISFNDKLNKNYISAGYTEYN 974

RESULT 15
Q81T21 PRELIMINARY; PRT; 2227 AA.
ID Q81T21
AC Q81T21;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF11_0354;
OS Plasmodium falciparum (Isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung B., White O., Brittan M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Sub B., Peterson J., Angioli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdey A.B.,
RA Martin D.M., Faircland A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

```

RA Fraser C.M., Bartell B.:
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum."
 RL Nature 419:498-511 (2002).
 DR EMBL: AB014841; AAN35938.1; -.
 DR HSSP: 0931C2; 1MWT.
 KW Hypothetical protein.
 SQ SEQUENCE 2227 AA; 262841 MW; 6D5D4BEEF643339C CRC64;

Query Match 6.3%; Score 252.5; DB 2; Length 2227;
 Best Local Similarity 19.1%; Pred. No. 0.024;
 Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;

QY 10 FKNLNGKE-----GSLKDTTGVH-----HQENESIXEK 42
 DB 630 YNNNNNNKCTTSINIKSENNKYPNNKSHVNSYMKVTHLPHRNAITSNNRNNEEYEXEK 749
 QY 43 S6FTIDRNIST-----IRDFENK---DKKLI-----KKKREVDFT-----SE 79
 DB 750 EK--DRNITNGNNVLYEYNNSCIPPLPKKMIPIDGVRNKSINKLNVNTQTSTSSVSY 806
 QY 80 TGRMEEDY-----KYDDKNIIVADDTLEETEKLDEIKSIYGLVSPSK 128
 DB 807 TKNIDENSFDMPIINGIRRESKISNNNNI---NGNSIGFNSKLDNYHQSMNV----- 858
 QY 129 DGHFELIGKISVSNKAKVYGNNTYSIETKATKYPFHSKTMTP-----DIYANIN-- 179
 DB 859 -----NESYPLKMMKQNNY-----IEHYDDKNNIPLVKNYEDTYSNIHNGI 900
 QY 180 -----DIVDGLAFAGDMR-----LFVKDN-----DOKK----- 202
 DB 901 HENSMIKYNNLKKACTFHYGSRNHQNNTEENLNIHQNNYSHYHNNGTVLKPLVNTNN 960
 QY 203 -----ABIKIMPEKIKETKS---EYPVSSYGNVIELEGDLKSKKPDNLTKMESGKI 253
 DB 961 VAVNEFADINLSAQKRIHSLKSMGYEDKEMENYRNKTIYNNNNNNNNNDN-----NI 1013
 QY 254 YSDSEKQY---LLKDNITLRKGYAL-----KVTIYNPEKTDMLBGNGVYSKEDIK--- 302
 DB 1014 YNDNEYCQYNNSYCFDHSDLKMMFPLNHQNSKLTLSNNKNSFFNGINVESKHLANPEI 1073
 QY 303 -----IQKANP-----N 309
 DB 1074 KTFANNSYPIINQGLINCPLOCLGYDSNQNNKANVYIKKNEYLANKIGIINVLKREG 1133
 QY 310 LRALS-----ETIYADSRNVEDGRSTOSVLSALDGFNIIRYOVFTFKAND 356
 DB 1134 LRKISTHNGKPFESFSGMDKNNVYMEGLNTQDNNNNNNKESC-DNIKHKRTKSLNFVSR 1192
 QY 357 K-GE--AIDKGNLYTDSKLVLPFGK-DDKEYTGEDKFNVEAIKED-GSMLFID--TKP 408
 DB 1193 SYGEHKSILVYOECYVKNKKLI--NKVNDKKY--EDNNNSYLNEDDNASMQFYEEETNSNP 1248
 QY 409 VNLAMD---KYNFNSKSNKIYVRNDEFLRGKISDKGFMNE-----LRVNESV----- 455
 DB 1249 YIVDQBNMKNNYV---NVLNNNSNYVDSKNTYDKSKENAKENKSDTLNNEENTHTLKD 1304
 QY 456 -----VDNYLLYGDHLIDNTRDFNLIKLVNDGDINDM--GKKDYKANGFPDKV--TDMG 506
 DB 1305 QKKKIQNNNEFISEQADIENINSQOEVEYEKEHEPL-WVINASNEKKSVEELIYSDMS 1363
 QY 507 NVYLTGTISDLAKAVGVAHQFLYDNVKEPVNIDPKGNTSIEYADGKSVFENINDKRNNG 566
 DB 1364 NRVTKRKYSDMNNVEVLNEDMLLTTEKYKQLEKENKIMDYETVEENINTIKENTND 1423
 QY 567 FPGELIOEHIYINGKEYSFNDIKQIDITLNIKIYVNDPANT--TVK--EPIINKONG 622
 DB 1424 INEEVR-----NEQGRESININD-----TWINHIDEPNDYTFIDIECVHNNEEN 1472
 QY 623 EYSELKPHRVYTIQNGKEMSSTIYSEBDFILPVYKGELEKGYQFPGWEISGFEGEKDAG 682
 DB 1473 MNSIEQYTFYHDTNRNHLVDK--NNQNF-----FEEGLNEALNFEKK--V 1516

QY 683 YVINISKD-----TFIKPVFKIIEKKERENKPTF-----DYSKKRON 720
 DB 1517 YIENNTKDXDHKGDSKTSNLTSLRNTICSENDHNEKNENTYVVRKGEKIRKYSMKKRN 1576
 QY 721 PQVN-HSQLN-----ESHKKEDLQREBHSQKS 746
 DB 1577 EKLNEENIYNNIYDKMDVHRQNDITKENDEEN 1609

Search completed: February 16, 2005, 19:19:52
 Job time : 188 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:11:30 / Search time 44 Seconds
(without alignments)
1690.354 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGEIASEKFKVLGNGKEGS.....ATVLDKNNISKSTNNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	2140	2 F95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase,
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypoch
4	250	6.2	1650	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T18427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhodopsin protein -
7	230	5.7	2401	2 T28676	rhodopsin protein -
8	228	5.7	4550	2 T18440	hypothetical prote
9	221.5	5.5	1711	2 T18429	hypothetical prote
10	219	5.4	1640	2 A24594	probable major sur
11	218	5.4	1631	1 SAZQK1	major merocyste su
12	218	5.4	1639	1 S05603	reticulocyte bindi
13	217.5	5.4	1546	2 G90603	reticulocyte bindi
14	215.5	5.4	1252	2 B42771	DNA-directed RNA p
15	214	5.3	2339	2 A45597	membrane nucleas
16	212.5	5.3	1125	2 E90598	hypothetical prote
17	212.5	5.3	2485	2 G70163	serine/threonine-s
18	212.5	5.3	2485	1 H71621	protein with 5'-3'
19	212	5.3	1188	2 A71621	microbial collagen
20	211.5	5.3	1104	1 A36866	REBA-H3 antigen pr
21	209	5.2	1558	2 B71603	hypothetical prote
22	207.5	5.2	4688	2 F82885	chromosome segrega
23	207	5.1	1141	2 B89824	hypothetical prote
24	207	5.1	1191	2 B97116	probable cell surf
25	207	5.1	1622	2 C71622	mature-parasite-in
26	206.5	5.1	1622	2 A81717	RAD2 endonuclease
27	206	5.1	1526	2 A45605	hypothetical prote
28	204.5	5.1	1516	2 E71619	
29	204.5	5.1	2500	2 G71609	

30	204	5.1	1245	2 D71613	GAF domain protein
31	203.5	5.1	3394	2 T18501	hypothetical prote
32	203	5.0	1166	2 T28680	fibrogen-binding
33	203	5.0	4981	2 T18489	hypothetical prote
34	202.5	5.0	763	2 G97026	superfamily 1 DNA
35	202.5	5.0	1939	2 T18372	repeat organellar
36	201	5.0	1712	2 C71618	hypothetical prote
37	199	4.9	888	2 E71608	ATP-dept. acyl-CoA
38	199	4.9	980	2 E71606	hypothetical prote
39	197	4.9	1286	2 AC1533	surface protein (L
40	196.5	4.9	2523	2 T18477	hypothetical prote
41	196	4.9	1714	2 E71609	Ser/Thr protein Ki
42	195.5	4.9	2380	2 E71604	hypothetical prote
43	195	4.8	1008	2 T18508	hypothetical prote
44	195	4.8	1302	2 JG5009	surface-located me
45	195	4.8	1365	2 T30822	Imp1 protein - Myc

ALIGNMENTS

RESULT 1

F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 09-Jul-2004
C:Accession: F95074
R:Retellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hel-
son, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Kadane, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MIMD:21357209; PMID:11463916
A:Accession: F95074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2140 <KUN>
A:Cross-references: UNIPROT:Q97RV6; GB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0641

Query Match	Score	4026	DB 2	Length	2140
Best Local Similarity	100.0%				
Matches	773				
	Conservative	0	Mismatches	0	Indels
					Gaps
QY	1	KLGEIASEKFKVLGNGKEGSLKDDTTGVVHHQENESIKKSSFTIDNISTIRDFENK	60		
DB	1334	KLGEIASEKFKVLGNGKEGSLKDDTTGVVHHQENESIKKSSFTIDNISTIRDFENK	1393		
QY	61	DLKLLIKKKFRVDDFTSTGKRMERYKYDDKGNITAYDDGTLEVTGLDEIKSKT	120		
DB	1394	DLKLLIKKKFRVDDFTSTGKRMERYKYDDKGNITAYDDGTLEVTGLDEIKSKT	1453		
QY	121	YGVLSPSKDGHEIIGKISNVSKNAKVYGNKYKSEIKATKYDFSKTMTFDLYANIND	180		
DB	1454	YGVLSPSKDGHEIIGKISNVSKNAKVYGNKYKSEIKATKYDFSKTMTFDLYANIND	1513		
QY	181	IYDGLAFAGDMFLFYKNDQKKAELIKRMEKIKETKSEYPVSSYGVNIELEGEDLSKN	240		
DB	1514	IYDGLAFAGDMFLFYKNDQKKAELIKRMEKIKETKSEYPVSSYGVNIELEGEDLSKN	1573		
QY	241	KEDNLTAKMSGKIYDSKQVLLDNITLRGVALKVTTPNGKTDMEGNGVSKKEI	300		
DB	1574	KEDNLTAKMSGKIYDSKQVLLDNITLRGVALKVTTPNGKTDMEGNGVSKKEI	1633		
QY	301	AKIQKAPVLRALSETTITADSRNVEDGSGTOSVLSALDGNITIRYQVFTFRKMDKGBA	360		
DB	1634	AKIQKAPVLRALSETTITADSRNVEDGSGTOSVLSALDGNITIRYQVFTFRKMDKGBA	1693		
QY	361	IKDKGNLVTDSSKLVLPKGDKEYTEGDKFNVAIKEDGSMFLPDTKPVNLSMDKNYFNP	420		

Db 1694 IDKGNLVTDSKYLFGKDDKEVTEGDKFNVEAIKEDGSMFLPITTKPVNLSMDKNYRNP 1753
Qy 421 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHIDNTDHFNIKLVNK 480
Db 1754 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHIDNTDHFNIKLVNK 1813
Qy 461 DGDIMDMCKDYKANGFPDXTDMDGNVYLQTVGSDLNAAKAVGVYQFLYDVKREVNID 540
Db 1814 DGDIMDMCKDYKANGFPDXTDMDGNVYLQTVGSDLNAAKAVGVYQFLYDVKREVNID 1873
Qy 541 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHOIYNGKEYTSFNDIKOIIDTFLNLIK 600
Db 1874 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHOIYNGKEYTSFNDIKOIIDTFLNLIK 1933
Qy 601 IVKDPFANTTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1934 IVKDPFANTTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1993
Qy 661 LEKGYPFGWEISGEGKKDAGVYNLSKDTFIRPVFKIEEKEENKPTFDVSKKDN 720
Db 1994 LEKGYPFGWEISGEGKKDAGVYNLSKDTFIRPVFKIEEKEENKPTFDVSKKDN 2053
Qy 721 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISKSSTNNPNK 773
Db 2054 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISKSSTNNPNK 2106

RESULT 2

A97942
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: A97942
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; Dehoff, B.S.; H
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaekunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97942; MUID:21429245; PMID:11544234
A:Accession: A97942
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2144 <KIR>
A:Cross-references: UNIPROT:O8DQF7, GB:A8007317, PIDD:AAK9365.1, PID:g15458138, GSPDB:C
A:Genetics:
A:Gene: prtA
C:Keywords: hydrolase; serine proteinase

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best Local Similarity 92.4%; Pred. No. 6,4e-151;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLGEIAEKFNKLGKESGLKQDTTGVHHQENESIKESPTIDRNIISTIDPFNK 60
Db 1339 KLGEISEDKLAKASPEENT-NNNAQKDESDKPDKSSVSGASLEIKTISTIDPFNK 1397
Qy 61 DLKLIKKKPREVDDFTSETGKMEYKYDOKNIIAYDDGTDLBYETEKDLDEIKSKI 120
Db 1398 DLKLIKKKPREVDDFTSETGKRIREYDYKYDDKNIIAYDDGSLQVETEFDEIKSKI 1457
Qy 121 YGVSPSPDGHFEILGKISNVSKAKVYGNVYSIKETIKATYDHSKTMPTDLYANIND 180
Db 1458 YGVSPSPDGHFEILGKISNVSKAKVYGNVYSIKETIKATYDHSKTMPTDLYANIND 1517
Qy 181 IVDGLAFAGDMRLFYKNDQKALIKIRNPEKIKETSEYVSVYGVNIEIGEGDLSEN 240
Db 1518 IVDGLAFAGDMRLFYKNDQKALIKIRNPEKIKETSEYVSVYGVNIEIGEGDLSEN 1577
Qy 241 KPDLNLTAKESGKIYDSSEKQOYLKDNIIILKRGVALKTYTYPNGKTDMLKNGVYSKEDI 300
Db 1578 KPDLNLTAKESGKIYDSSEKQOYLKDNIIILKRGVALKTYTYPNGKTDMLKNGVYSKEDI 1637

Qy 301 AKIQANPNLALSTTTIYADSRNVDEGRSTQSVLMSLDGNNIRYVFFPKANDKGBA 360
Db 1638 AKIQANPNLALSTTTIYADSRNVDEGRSTQSVLMSLDGNNIRYVFFPKANDKGBA 1697
Qy 361 IDKGNLVTDSKYLFGKDDKEVTEGDKFNVEAIKEDGSMFLPITTKPVNLSMDKNYRNP 420
Db 1698 IDKGNLVTDSKYLFGKDDKEVTEGDKFNVEAIKEDGSMFLPITTKPVNLSMDKNYRNP 1757
Qy 421 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHIDNTDHFNIKLVNK 480
Db 1758 SKSNKIYVRNBEFYLRGKISDKGFWNMLRVNESVVDNYLIYGDHIDNTDHFNIKLVNK 1817
Qy 461 DGDIMDMCKDYKANGFPDXTDMDGNVYLQTVGSDLNAAKAVGVYQFLYDVKREVNID 540
Db 1818 DGDIMDMCKDYKANGFPDXTDMDGNVYLQTVGSDLNAAKAVGVYQFLYDVKREVNID 1877
Qy 541 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHOIYNGKEYTSFNDIKOIIDTFLNLIK 600
Db 1878 PKGNISIEYADGKSVFVFNINDKRNNGFDEIOEHOIYNGKEYTSFNDIKOIIDTFLNLIK 1937
Qy 601 IVKDPFANTTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1938 IVKDPFANTTVKEFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1997
Qy 661 LEKGYPFGWEISGEGKKDAGVYNLSKDTFIRPVFKIEEKEENKPTFDVSKKDN 720
Db 1998 LEKGYPFGWEISGEGKKDAGVYNLSKDTFIRPVFKIEEKEENKPTFDVSKKDN 2057
Qy 721 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISKSSTNNPNK 773
Db 2058 POWNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISKSSTNNPNK 2110

RESULT 3

T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T28317
J:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A>Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: UNIPROT:Q9YVT6; EMBL:AF063866; NID:g4049647; PIDD:AAK97677.1; PID:g4
C:Genetics:
A:Note: MSV156

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0015;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;
Qy 7 ESKRPNLNGKSGSLKQDTTGVHHQENESIK-ESKSTIID-RNISTTRDENKOLK 63
Db 196 EIEFPNIDNVKKEINKQDEINKLDESKKEFIRKQEBELNITPKQEBELIKKNDKEIN 255
Qy 64 KLIKKKPREVDDFTSETGKME-----EYDYKYDDKGNIIAYDDGT--DLEYTE 111
Db 256 FNIDKQKLLQINSKINTLNEITKGMVNLTEETKKNKISNLQNELINKOSTYKSLDEKQK 315
Qy 112 KLDEIKSKIYGLSPSKDGHFEILGKISNVSK-----NAKVYGNVYSIKETIKAT 161
Db 316 LLDDELKNNINNTSLYKNSN-----TKITNIQQLLESSITDPMNANI-----NINELKSKIX 367
Qy 162 KYDHSKTMPTDLYANINDIYDGLAFAGDMFLFYK--DNDQKA-ETIRNPEKIKET-- 216
Db 368 LFDNDIQKLNNDITEONNKITD--FFNNSTRIPIFEKIDTEYKAKIIDIDIGNNMLQKLEESYK 425
Qy 217 -----KSEY---PYVSVGNVIEIGEGDLSEKPKPNLTGMEGKKIYDSSEKQOYLKDNII 269

```

Db      426 KIDQTEYYKKNKINKEYNDIIEIKNNNLQKLEENKKIDQTEYYKKNKINKEY--NDILE 483
Qy      270 LKRGVALKVTYTPG-----KIDMLENGVYSKEDIKQKPNRLALSETTYAD 321
Db      484 LKNNNLQKLEENKKNINDKLTLYKNDIESNTBELFKLNTISDFKDSREIAKLN--TEYEQ 541
Qy      322 SRN--VEDGRSTQSVL-----MSALDGFNIIRYQVFTFRMNDKGEAIDKQNLVTDSK 373
Db      542 LKRDLEENKNTKMLKLSDNKLSLE-----QYDSKKNIL--DGIDKIVNSLEKKN- 592
Qy      374 LVLFGKDDKYEYGEDKFNVALEKED--GSMFLIDTPKPNVLSMDKNYFN--PSKSNKTY 427
Db      593 ---DKIDEFNSIEKFIDYVNIENKFINLDSIIINKIINNDFKEXYINSKIDSKSNES 648
Qy      428 VANPEFY-----LRGKISDKGFMWELRVNESVVDYLLYGLDHLIDNTED 472
Db      649 TWFDDIFNAKQIATITNNIENISNKIKD-----LNEFTISNEDSSKEL--LDEIRK 698
Qy      473 FNIKLVKDDGIDIMDMKDYKANGFPDKYTDMDGNY--YLQGYSDLNAKA-----V 522
Db      699 YKQGPD-KIKDAMNTEVKSFE--NTLQKIDISIKSNINELNADIIINTKANDLDDKLNKY 756
Qy      523 GHYQFLYNNKPE-----EVNIDPKGNTSIEYAD--GKSVYFNINDKRNNGFDGEIOEQHI 576
Db      757 GSEFKLVNNSASDLDTIQKNDDEKVKOLNEYLEKKNKQSIIEINDIVN----- 805
Qy      577 YINGKEYSFENDIKOIIDTKINKIYVVDPAFVTVKEPILNKDQGEVSE--LKPHRYT 633
Db      806 FY--KELIFKNTE--TNKSLN-ELLTND--DINDKIFLYKELNKISTNNLKIYK-- 855
Qy      634 VTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGMEISGFEKKDAGVNLKSCOTF- 692
Db      856 NEIDNVNEKLSIYENLQFINSLFSLIEFNQ-----SITSH-----INFLNTLA 900
Qy      693 -IKPVFKLIEKKEENKPTFDVSKKKNPQVNSQLNESHKREDLQREHESQKSTQD 751
Db      901 GINDVNLKNTLKIIMDTTRRGDTNIR-----DEIKNOISSENIKSOKEFNEKN--XD 950
Qy      752 VTATV-----LDKNISKSSTNN 770
Db      951 LKKLISFNDKLNKNISAGTYEN 974

RESULT 4
T18444
Hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18444
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1650 <LAW>
A:Cross-references: UNIPROT:O77328; EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CAB111
A:Map position: 3
A:introns: 1597/3; 1625/3
A:Note: C0385c

Query Match      6.2%; Score 250; DB 2; Length 1650;
Best Local Similarity 20.8%; Pred. No. 0.0034;
Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;
Qy      20 SLKKDITGVYEHQNEESIKESSTIDRINISTIRDFENKDKLKKKKFRE--VDPF 76
Db      394 SLWENNEKRIYKEIKESVKNQEV-----HFKND--DSGIKKKNSSECLDEG 443
Qy      77 TSETGR--MEYDYKYDKGNITAY--DDGTDLE--YETKLDIKSKIKYIVLSPSKD 129
Db      444 KKTYKYTIIIEQRYVFNDRNNNAVYIKDTHKKEKGYILNMIYQSEYKYG--SNMK 501

```

```

Qy      130 GHFEILGKIS-----NVEKNAKVVYGNVYKSI--EIKATKYDFHSHKWTFDLYANINDV 182
Db      502 DEMETIYQHTNDFNINENLANKIYF--DYEGDPEPKKKKKLDLHDHYQKEXKANNINDL 560
Qy      183 DGLAPAGDRLFVKNDQKKAIEKIRMEPKI--ETKSEYPVYSSYGNVIELEG-- 235
Db      561 K-----DHLINDKETKEKNEIEEEKKNKIEIEEEKKNKIEIEEEKKNKIEIEEEK 614
Qy      236 ---DLSKPKPDLTQMESGKIYSDEKQYLLKNDI-----LR 271
Db      615 KKEIEEEKKKKI--EMEEKKNKIDKKQNTYANDKIISHIDVNCNIKIDALLDHIIEEK 673
Qy      272 KGVALKVTYTPGKT-----MLENGVYSKEDIK--IQKPNL-- 310
Db      674 KTGHEINLYKEIKRYEQKMLNDENSIMLEHKKNTHTQVNNNLDDTKMLQKELKILN 733
Qy      311 RALSETTYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFRMNDKGEAIDKQNLVTD 370
Db      734 NDKKTFILSKSNI-----TSNVLSSKIPG-----TLSTKGLNATIKTIKKD--VTD 779
Qy      371 SKGLVLPF-----KODKE--YTGEDK--NYEA-----IKEDSMLFIDTK 407
Db      780 NEKKKYVHDHRKONITIKNKEFINIYKGRNANVEIGSEVCNNKVNAGDNRKMMVENK 839
Qy      408 P-----VNLSDKNYF-----NPSKSNKIYYANPEF 433
Db      840 QGDNNMNVENKQGBENNVIVANIERSSSFFTHHRKNITSTSDTKAKNEQIKYPHL 899
Qy      434 YLRGKISDKGFMWELR--VNESVDVNY-----LIYGLDLDN----- 469
Db      900 Y--SNKNEDEKDKSIFLKININENIKKQYKDEKISTLEKKVFVKDQNVITNNDDEKHTSSK 958
Qy      470 -TRDENI-----KLN----- 478
Db      959 INDDFNITTDKQKGLNNPVDLNRKIKETKILEKDSHMSKIONNIMKKTTPPTNK 1018
Qy      479 -VKDGDIMDMGKKDYKANGFPDKYTDMDGNVYL-----QTG 513
Db      1019 GISSSTSISSSTKRNKDCGIIERKNLSLKYTCIRKNQNVDSIKLNDKADLYKDKKTS 1078
Qy      514 YSDLNKAVGAYHAYQFLYDNVKEPVNIDPKGNTSIEYADGKSVFN--INDKRNNGFDG 569
Db      1079 FNDINRAKGNKFK-----KADY--PKNNMVDYTNKGR--VFNVPVTLNNYRNN----- 1124
Qy      570 EIQEQHIYINGKEYSFENDIKOIIDTKINKIYVVDPAFVTVKEPILNKDQGEVSELKP 629
Db      1125 -----YIRSNK--NNVK-----NGKRV-----GIKKIVLKE--KOKSLHP 1156
Qy      630 HRVTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGMEISGFEKKDAGVNLISLK 689
Db      1157 EGVEADKILNSYNDKYLEKQFPDIIINEBEK-----YKNNK--KKYK--K 1201
Qy      690 DTFIKPVFKLIEKKEENKPTFDVSKKKNPQVNSQLNESHKREDLQRE--HSQKS 746
Db      1202 SNSIPIIKIERKSDNDN--DNIKNDNINSNNINSNDKCLFLSKEDRVHLKNN 1258
Qy      747 DSTKQVTAIVLDKNNISKSSTNNPNK 773
Db      1259 NITVNTMMFRKQNSCDNNTTSLNKK 1285

RESULT 5
T18427
Hypothetical protein C033c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ

```

A.Molecule type: DNA
 A.Residues: 1-3724 <LAN>
 A.Cross-references: UNIPROT:O77320; EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB111
 C.Genetics:
 A.Introns: 307/1; 1545/2
 A.Note: C0335C

Query Match 5.8%; Score 235.5; DB 2; Length 3724;
 Best Local Similarity 18.7%; Pred. No. 0.041;
 Matches 192; Conservative 150; Mismatches 312; Indels 371; Gaps 45;

```

QY 5 IAESFKVLGNKEGS--LKK-----DTGVHHHNGEESTK-----EKSPT 46
DB 249 IPKNLFENLNKKQHNDYQNRNIIIMDVNDINPLHPBQONSLNKKKCLTGTNKEKYM 308
QY 47 I-----DRNISTIRFENDKLKLKKFREVDFTSETKMEEDYKXDDGK 95
DB 309 IPKNGEYDQKEKESLITINQNDKKYKKK--KKKXSELQDSNINNTLTSKRYTT--- 363
QY 96 NIAYDGTDLLEYFEKLDEIKSKYGVLSPSKDGHEILGKI SNVSKAKYVYGNVYKS 155
DB 364 -----CGMDKETETDEQNKR-----RNTEVASICVNCVETEKOKRKTNNKK 407
QY 156 IEIKATKYDFHSHKMTFPLNYANINDIYOGIAPAGDMR--LFGKNDQKKAELIKIMP--EK 212
DB 408 ---SVQKODLVEENALD---NL-IITDGINFDVTKECKITIDNNNNENIDNNIYVENNK 460
QY 213 IKETSEYVYVSSYG-NVI-----ELGEGDLSKNKPNMLTKMESGK 252
DB 461 LKDDQS-YDLFSSEGNLILGVNBESEFNEFENIEKLGQRKQBKNKNDK-----T 512
QY 253 IYSDSEKQOYLKKNIIILKGYALKVTTYNPGKTMLEBNGVYSKEDIAKIQKAN-PNLR 311
DB 513 IYNNNEBQTDILNRNI-----NKIESINNNDN---NNNNNNKKEFKIRTEHILNKE 562
QY 312 ALSETTIYADSRNVDGSGTQSVLMSALDGNIIIRYQVTFPFGNKGALIDGMILYDS 371
DB 563 SISKAIKSSPSKDKK-----IKKLTYTNQEDSTELKKELEIITNN 604
QY 372 SKLVFGKDDKEYTGEDKENVEAIKEDGMLFIDTKPVNLSDKNYFNPSKN----- 424
DB 605 KNNVY--BEDIIGSNBDEYIHVKE-----NLKEDABENNDKKNKNNKTKE 650
QY 425 ---KIYRN-----PEFLRGKISDKGSENNMELRVN--ESVDNLYLIGDLHIDNTRD 472
DB 651 ILKSNKYLENKRTELEELRGK--NNIFKDEKNSLGEVYINIEIQINEBNKINDIOD 707
QY 473 FNIR-----LNVKD---GDIWDGMQDYKANGFPDVTDM----- 505
DB 708 GNISKOKIIQSSSRNTDFTNIDKISLNDLLEKKEKRGKSQHFIDMLVKADKNEISENINK 767
QY 506 -----GNVYLQT-----GYSDLNKAAGVHY----- 526
DB 768 ICDNNINNIYDESINNIYDESINNIYDESINNIYDESINNIYDENINNIYDENINNIYDE 827
QY 527 -----QFLYDNKPEVNID-----PKGNTSIF-----Y 549
DB 828 NINNIIYDEGINIKCDNIIENKNIKTNDIYQVEBENNESIEKNELMISLAKDINNYYNF 887
QY 550 ADGKSVFN-----INDKRNNGPDEIOE----- 573
DB 888 KENVDFIPIKIRGEBLSLKIDKNIKONNDDDEYIMDNFYENDTIINHKHEITNKEIDPLEI 947
QY 574 -----OHIIYNGKEYTS--FNDIKQIIDKTNLKIVVDPFARTTYKEPILNKDGE 623
DB 948 NTQNEFIERLIDIKKKKYTYNDHFPADKKFYE-NM-KILNKMCKKNKQDEFPKTDGETGS 1005
QY 624 VSELK-----PHRYVTYTIQNGKMSSTIYSE 649
DB 1006 LOSHKIKKKYKKEKRGKDNBEEKNLYDENQVSVLYSDHKIEDIODIHSIQNTICDE 1065
QY 650 EEFILPVYVGELEKGYQDFDGEWISGPEGKDGAGYINLSKDFIFKPVFKIEKKKEBENK 709
  
```

A.Molecule type: DNA
 A.Residues: 1-2269 <KEB>
 A.Cross-references: UNIPROT:Q26223; EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304
 C.Species: Plasmodium yoelii
 C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C.Accession: T28677; C45521
 R.Keen, J.; Siba, K.; Brown, K.; Holder, A.
 Mol. Biochem. Parasitol. 65, 171-177, 1994
 A.Title: A gene coding for a high molecular mass rhotyry protein of Plasmodium yoelii.
 A.Reference number: Z20508; MUID:95021522; PMID:7935623
 A.Accession: T28677
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-2269 <KEB>
 A.Cross-references: UNIPROT:Q26223; EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304
 R.Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
 Mol. Biochem. Parasitol. 42, 241-246, 1990
 A.Title: Identification of the gene for a Plasmodium yoelii rhotyry protein. Multiple cc
 A.Reference number: A45521; MUID:91101660; PMID:2270106
 A.Accession: C45521
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 2131-2269 <KE2>
 A.Cross-references: GB:M34283

Query Match 5.8%; Score 231.5; DB 2; Length 2269;
 Best Local Similarity 19.7%; Pred. No. 0.032;
 Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 37;

```

QY 37 ESIRKESFTIDRNSTRDPENK-DLKKLKKKPREVDFTSETKMEEDY----- 87
DB 109 BGLKLELNKIKDIYAKIEYKAYVLELGEIEKNNAVYIDELANGSPRYVTGIEKNKTY 168
QY 88 -----DYRYDDKGNIIAYDGTDELEYETEKLEIKSKIYGVLSPSKDGHF 132
DB 169 NTIKSYQPIYEGDIDTFYNELSLIVKEDPIDIEDKT-KLENLRSKIDYVDKIQKEI 227
QY 133 E-ILKISNVSKNAKVYGNVYKSEIEIKATKYDFHSHKMTFPLNYANINDIYDGLAPADM 191
DB 228 ETVKSHLNNIETNNKL-----PNTILEIKKYIYDEISK---ELNKMLEDFFNRK--EKEL 276
QY 192 RLFFVNDQOK--AEIKIRME-----KIKETSEYVYVSSYGVNIEGSE 234
DB 277 SNKISDYKKGQOLSEYKSLERHNSQTNVDNTEBBAKQVYDGSNHEMTIIPNE 336
QY 235 GDLK-----NKEDNLTQMS-----GKIYSD--SEKQOYL-LKONI---ILKGYAL 276
DB 337 DEISKIIEVKTMDIELSKVNTYIDFNKRYKETVNSHSGFTLETDKIKAEVSDKELK 396
QY 277 KATTYNPKTMDLEBNGVYSKE--DIAKIQANPRLALSET-----TLYAD-- 321
DB 397 CEQSFNDNRKSLINERKNSIEKEYOINTLKQVDEYIKVCKSYKESITPSSKQITLKML 456
QY 322 SRNVDEGSGTQSVLMSALDGFNIIRYQVFT-----FKNN-----DKGEAI----- 361
DB 457 NQNTKYVETNSIDSYIEKEF-----QILTQKQTLKPKFTBFSLNHNEANNELIKYFS 512
QY 362 DKDGNLVTDSSKVLFGDDKEYTGEDKENVEAIKEDGMLFIDTKPVNLSDKNYFNPS 421
DB 513 DLKANLGINENMNLVNGTEKEKTFND-----IKENIHIINEISLEIKIKHAIYVNS 566
QY 422 KSNKIYVNPFIYLRGKISDKGSENNMELRVNESVDNLYLIGDLHIDNTRDPNITLANK- 480
  
```

```

Db      567 ESTERBIGINISLTKVFEK-----VKENVTLANKIKETLKHYPDSFGKEGNIKY 618
Qy      481 -----DGDINDMGK--DYKANGFPD-----KTDMDGNV-----YLQGYSDLANK 520
Db      619 TDKIKKINDIDNAVSQDIOHINGLDDIQKSESYSVEMEQINLKEVSNFEISDNVE 678
Qy      521 AVGVHQPFLYDVKPEVNI-----DPKNTSIEYADGKSVFNIN-----560
Db      679 GKXKQOIIIVTKIDKKKNLYEELINKLSIEKESKONTSLKVKDINLSYQNLGWLFILE 738
Qy      561 --DKRNGFDEIIOEHYIN-----GKEYTSFNDIKQIIDKTLN- IKI-----601
Db      739 QIDEEKKKAENITKSMEAYIIDLDNKKKSQEIETEMDIKMDINKEMELAKISHDDKCC 798
Qy      602 -----VKDFARNTVAFILNKDTGSELSKPRVYVTTIONGKE 641
Db      799 HDKSNKHENISIDYDKSSKIIDFRESKDIND-INNKLOKNVSEQNHNSDIN-----851
Qy      642 MSSTVSEDFILPVYKGELEKGYQFDGWEISGFEGEKDAGVYINLSKDTFIKPVFKIE 701
Db      852 -----QCLNVANINYLKLNKIKIIDKVK 877
Qy      702 EKKEENKPTPVSKKKDNPQVNHSQLNESHREKDIQREHSQKSDST---KDV-----T 753
Db      878 EYTSIEK-----NKKINDELNNSEKVIKIEGDLSLKESCKSKINSTLDDKDIDECIRK 932
Qy      754 ATYLDKNNISKSSTTN 770
Db      933 INVLLKNILNETNITN 949

```

RESULT 7

```

rhodtry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676, A45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A>Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507, PMID:97077455, PMID:8920022
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q26216; EMBL:U36927; NID:91041784; PID:91041785; PIDN:AA8412
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A>Title: Identification of the gene for a Plasmodium yoelii rhodtry protein. Multiple c
A:Reference number: A45521, PMID:91101660, PMID:2270106
A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

```

Query Match 5.7%; Score 230; DB 2; Length 2401;

Best Local Similarity 21.0%; Pred. No. 0.04; Matches 187; Conservative 138; Mismatches 292; Indels 274; Gaps 46;

```

Qy      2 LGBIASEKPNLNGKSGSLKQDTTGV--HHQENESIKESGFTIDRNISTIRDFENK 60
Db      589 IDKIYDKFENITDKTELEFRTGLSNHESNN---KELTYTYDLKANKGKXKEN- 644
Qy      61 DLKLLIKKKFREYVDFTSETGKMEBYDYKDDKNIIA-----YDGGTDLEVEY 110
Db      645 ---MLYKQFNKSG-----KAVEDIKKKNVDINKIVSNIETIYTSIYNEDTENEL 693
Qy      111 EK-----LDEISKIYGVLSPEKDGHEFLIGKISNVSRAKATYYGN 151
Db      694 GKSIELLNTKVLKVKANTNINLEIKEL-----KDYDFQDFGK---EKNIKYPDEN 742

```

```

Qy      152 NYKS-IEIKATYDPSHKTMTEFDLYAN---INDI---VDGLAFAGDKMLFYKNDQKKA 203
Db      743 KIKNDIDTLNQKIDKSIETLF-EIKNSENHIDIKQDILKVKVFNKTMF--NEDPKXI 799
Qy      204 EIKI-RMPEKIKETSEPEYVSYSYGNVLELGGD---LSKNPDLTKNES-GKIY-----254
Db      800 EKKINIEYKIDKKNIYKIEDKLNEISKLINDKTSLEKKNINLSYKSGISGNIFLQOI 859
Qy      255 -SDSEKQOYLK-----DNI-----ILRKGYA 275
Db      860 DEKKKAETHITAMAAVYIDDLNIIKKKSQEIEMKNIMMDIMDIHKEMKALNISHDYK 919
Qy      276 LKVTI--YNGKTDMLENGVYSKEDIATQKAPNLPALSETTYADSRNVEDGRSTOS 333
Db      920 IYHTTSKNHEKISDIRKNSLKIIDFSESYSYNIIDKELEKNVLESQNNNTDINO---975
Qy      334 VLMSALDG-FNIIYQVLFPRKNDGKALDKQGNLVYTBSSKVLVFGKDKETGE-----387
Db      976 -YLSKIENTYINIL-----KLNKIKIKIDK-----VKEXTDEIEKKN 1010
Qy      388 DKFNYE-----AIKEDGSMFLIDTPKPNVLSMDKQYFNPSKSNKIYVRPEFYLRG 438
Db      1011 KKNBELNSSEKIITQLENSLSKECQSK-ISTIDNVSECIRK---ITLKYTYI---1063
Qy      439 ISDKGGFNWEKLVNESVVDNY-LIYGDHIDNTRDPNITKLVYKDDIDMDGMDKYKANGF 497
Db      1064 VNEKNINNTYFNAEBYQNVSLNFTNIEIMADTKQYI-LNIK-----KNNG- 1109
Qy      498 PDKVTDMGNVYLOQGYSDLNAKAVGHYQPLDYNVKPEVINIDPKGNTSIEYADGKSVYF 557
Db      1110 -----TNNTDYNIKELKEHK-----KSNVYDEAGKNTQEIKNKNEI-- 1147
Qy      558 NINDKKNNGFDGEIOEHYINNGKEYTSFNDIKQIIDKTLN-KIIVKDF--ARNTVVE 614
Db      1148 -----FEKYBOEVYVLN-KYYAV--ELKAKFDKTKYNSBOIIEIKIKAHNTFTSQ 1195
Qy      615 FLINDTGEVSELKPRVYVTTIONGKENSSTVSEBDFILPVYKGELEKGYQFDGWEISG 674
Db      1196 -ADSEKKKQNEIKNQEIIEDEVAKNKSN-----KALID-----1229
Qy      675 FEGKGDAGVYNLSKDTFIKPVFKIIE---KKEENKPTPVSKKKDNPQV-----723
Db      1230 -----IOLSYEPF-KIFELKIKIDRTSDCLKTKDIEKTSILSDTQETKLI 1278
Qy      724 -NHSOLNESHREKDIQREHSQKSDSTQVATVLDKNNISKSSTTNPNK 773
Db      1279 EKNKINLTLEKLESUKQKNIEDQKE-----LDEVANSKINIESNVNQ 1324

```

RESULT 8

hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18440

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18935

A:Accession: T18440

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4550 <LAW>

A:Cross-references: UNIPROT:O77336; EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CA811

A/Map position: 3

A/Note: C0425w

```

Query Match 5.7%; Score 228; DB 2; Length 4550;
Best Local Similarity 21.6%; Pred. No. 0.11;
Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;

```

```

Qy      5 IASEKPNLNGKSGSLKQDTTGV-----EHHQENESIKESGFTIDRNISTIRD 56

```

```

Db 746 IEESHSISGSHNEIYASIKCAQNGVRDNKLDSGHNDKDKDNKKD---DKDNDKKD 802
Qy 57 FENKDLKLLIKKKFREVDFTSETGKRMEEYDYKDDKNIIAYDDGTLEJE----- 109
Db 803 GPKDDDKKDDKDYDDGDD-----KYDDDNKTDDDDI---YDDDDPFNDHDSERSK 853
Qy 110 -----TEKLDEIK-SKIYGLSPSKDGHFELIGIS-NVSKNAKYYYG 150
Db 854 RLSNVDLLKDKKKKGSILKEEDTLCKSKM1NVOEKESKMYCMNTNNCNVYKE-KNYFV 912
Qy 151 NNYKSEIKATKYDFHSHKTMFTDLVANINDIYDGLAFAD-----MLF-----YKN 198
Db 913 NDYMH1INNLSHHEMINMKSPGINNLFKDVHFIHKDFINILIKLKEKIIQVLQN 972
Qy 199 DQKKAIEIKRMPEKIKETSEYRYVSSYGNV-----TELGEGLS-----KN 240
Db 973 NYKQVCVFCKTTGIR-TKCMFPCSTYFHIYCYNNKWHYRKKKNDQCSQKSKRN 1031
Qy 241 KPDNLTQKE--SGKIYSDSEKQOYLKDN1ILRKGYALKVTTYNPCKTMDLEGNVYSKE 298
Db 1032 QRAASVVRGAEKCKDKDGRSLDRKCDKYTRRSNKEYESANDSGKKV--NGDDKKKD 1089
Qy 299 DIAKIQKANPMLRALSETTIYADSRVDEGRSTQSVLMSALDGFNIIRYQVTFKMDK 358
Db 1090 NIGKNNNNNNNNDDSDSNNNGDSNNNG-----DGNNTINGDNNNNINGDN 1143
Qy 359 EAIDKQGNLVT-----DSKLVLFSGDKDEYTGEDKFN 392
Db 1144 NNINDGNNNNNNCBERKQDGNQKRRKRRKRRNNNDTSLSVLANKDSRQKKKNKYG- 1202
Qy 393 EAIKEDGSMFLFDTPVNLSPDKATFNPSPSKNKIYVRNPEYLRGKI-SDKGGFWELRV 451
Db 1203 ---KDH-----DOEKDPSFKINNRSKN-----KVKKDKKEVGN---I 1234
Qy 452 NESVVDNYLIYGLHIDNTRDFNIKLVNDGD-----IMDMGMDKYKANGFPKPTMDGN 507
Db 1235 NDDKINN-----NNDNNKNNNT--NGDNKNNNNINGDNKNNNNIN--DKI---NKN 1278
Qy 508 VYLQGYSDLNKAVGVHY-QELYDNVKEVNIIDPKGNTSIEYADGKSVFVINDK--- 562
Db 1279 INGDKINNINDKISKVHSHNYDN--EILTHEQNNFN-KNKHTKRRKRSINNKEMND 1335
Qy 563 ---RNGGPDG-----IQEHIYNGKEYTS-----FNDIKQIIDKTLNIIKIVKQFA 607
Db 1336 ILNNKNNDDKDKLNKNFLEMYOKIISDKITSISNNMKDIKNKD-----IKQIK 1387
Qy 608 RNTYKFEFLNKDGEVSELKPHRYVTIIONGKEMSTIVSEDFLIPYKGELEKGYQF 667
Db 1388 DIKNIKDI---KQIKDIKDIKDIKQIKVANKIKD-----IKSNED----- 1424
Qy 668 DGMELSGEGKKDAGYV-----INLSKDT--FIKPVFKIEE--KKEENKPTFVSKKK 718
Db 1425 ---IKDIDVNDVSVHLNKCVCNNKSDIQDFCPSDDKNINNIKKNNKINITDQY--- 1477
Qy 719 DNPQVNHSQLNSHKKEDIQREHSGKSD-STKDVATATVLDKONTSSKSTNNPN 772
Db 1478 -NPLVDSSSSSTDCNKYKKEKKAAYKIGMINKIKLHNNNNNNNNNNNNNNNN 1531

RESULT 9
T18429
Hypothetical protein C0345w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18429
R:Lawson, D.; Bowman, S.; Barrett, B.
Submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18429
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <1AM>
A:Cross-references: UNIPROT:O77322; EMBL:Z98547; NID:e1325376; PID:e1325381; PIDN:CAB111

```

```

C:Genetics:
A:Note: PF0345w
Query Match 5.5%; Score 221.5; DB 2; Length 1711;
Beet Local Similarity 19.7%; Pred. No. 0.05; indels 243; Gaps 37;
Matches 160; Conservative 134; Mismatches 274;
Qy 29 EHHQENESISEKSS-----FTIDRNISTIDFENKDLK-----KLKKKFEVD 74
Db 994 DNNNDNNNNNNYDSSSSNNHYIILTNDKRLN-MDNF1NNNLNNSQNKVLEKLEVIN 1042
Qy 75 DF-TSETGKRMEEYDYKDDKNIIAYDDGTLEJEYTEKLDEIKSKIYGLSPSKDGHPE 133
Db 1043 NVKLTQTS-----NYESGNTNNSKDE-----HNISDSKSKEDTJLNSKSSYE 1086
Qy 134 ILGKI--SNVSKNAKYYGNNYKSEIKATKYDFHSHKTMFTDLVANINDIYDGLAFAD 191
Db 1087 YNNKLIQSTSNKSLGAYENNLFSGKKKKK--GYLADI-RHINDIQD-KYREDL 1138
Qy 192 RL-----FVKNDQKKAIEIKRMPEKIKETSEYRYVSS--YGNVIELGEGLSKNRP 243
Db 1139 NINCYNKYVIEEBEKGHL-----LPEL-----EYVLVSDEKFG-----LNKIKND 1179
Qy 244 NLTKMESGKIYSDSEKQOYLKDN1ILRKGYALKVTTYNPCKTMDLEGNVYSKEDIKI 303
Db 1180 N-----NIYMKHQNYHNLVDN--QKQIILPDTNQ-----VSI 1212
Qy 304 OKANPMLRALSETTIYADSRVDEGRSTQSVLMSALDGFNIIRYQVTFKMDK--GEAI 361
Db 1213 QR-NNNNNVKIKTHIYEVKNNKDKR-----YDNFCDDKKKIYYNII 1255
Qy 362 DKDGNLVTDSSKLVLFQDKDEYTG-----EDKFNVEAIKEDGSMFLFDTPKVN- 411
Db 1256 NSDKDIYHNN--IYTKNEKEGIGNHILNRNDKQITFELLKLDGVKPEFLDTPKDSYTD 1312
Qy 412 ---SDKATFNPSPSKNKIYVRNPEYLRGKISDKGFNMELVNVDNYLIYGLHIDN 469
Db 1313 CHNKKENILNNTNKK-----EDHQIILDVADK-IFN---ETNM1TMMNNKIYDKNVAHE 1362
Qy 470 TRDFNIKLVNDGDIMDMGMDKYKANGFPDQKTYMD--GNYVLQGYSDLNKAVGVHYQ 527
Db 1363 KKCITNDVYHNNMDLSTISKNNENLFDIYQOKNRIGDIYMR----- 1407
Qy 528 FLYDNVKEVNIIDPKGNTSIEYADGKSVFVINDKRNNG-----FDGEIQEHIYNGKEY 583
Db 1408 ---INILQEDDDDDHNN-----HNNNNNNNNKILFXYTKNDQDLHNNKNL 1452
Qy 584 TSPNDIKQIIDKTLNIIKIVADPANTTVKPEFLNKDTGVEVSELKPHRYVTIIONGKEMS 643
Db 1453 EGTEFFSDPIEKNNKIKIKNNKESYHKLIDESILSNKNNKVSLL-----LINNNKSS 1505
Qy 644 STIVSEEDPFLPVYKGELEKGYQFGMEISGPEGKQDAGYINLSKOTFIKPVFKIEBK 703
Db 1506 SY-----DNNKNNNNKNNNNK-----NANNEN 1527
Qy 704 KEEENKPTFVSKKKDNPQVNHSQLNSHKKED-----IQREHSGKSDSTKDVATV 756
Db 1528 NNNKNNKNNNDSPSKDNNL1NNNDNNNNNNNDSPSKDNNL1NNNDNNNNNNNNKVIKKEI 1587
Qy 757 LD--KNNISSKST-----TNPN 772
Db 1588 IDDEKNDIHKRDNIYIKDVSPLINHPN 1618

RESULT 10
A24594
Probable major surface antigen (83k, 19k, 42k) precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jul-2004
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholas,
Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Plas

```

A:Reference number: A24594; MUID:86014355; PMID:2995820
 A:Accession: A24594
 A:Molecule type: DNA
 A:Residues: 1-1640 <HOU>
 A:Cross-references: UNIPROT:P04933
 C:Superfamily: major merozoite surface antigen
 C:Keywords: surface antigen

Query Match 5.4%; Score 219; DB 2; Length 1640;
 Best Local Similarity 18.6%; Pred. No. 0.071;
 Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

```

QY 23 KDTTG-VEHHHQBESIK-----EKSFTIDRNISTIRDPENDKLKKKKFREVD 75
DB 232 KNVGMEDEYIKKKKTIENINELIESKTTIDKKNATBEE-----KKKLQAOY 283
QY 76 FTSCTGRMEYDYKYDDKNITAYDGDGDLLEYETKLDKIKYIGVLSPSKOGHEILL 135
DB 284 DLSYVKOLEE-----AHNLISV-----LE---KRIDLK-----KNENIKELL 319
QY 136 GKISNVSKAAKYVYNNYSIEIKATKYDPH-----SKTTFPDIYANINDIVDGLAFA 188
DB 320 DKINELKPPPPANSNGTPTLTDKNKKIEHEKEIKETIAKTIFKNIDSLFTDPL----- 373
QY 189 GMRLEFVKNDQKAEIKIRMPKIKETSEYV-----YVSSYGNV-----IELGE-GDLRK 239
DB 374 -ELEYYLRKN-KNIDISAKVETKESTPEPEYNGVYPLSYNDINNALNELNSFGDLT- 430
QY 240 NKPDNLTKBESGKIYDSEKQOYL--LKONIIL-----RKGYALKTTYPGKTIDL 289
DB 431 -NPDYTKPSKNITYDNERKKFINIEIKIKIEKKIESDKSYDRSKSLN---DIT 485
QY 290 EG-----NGVYSKE-----DIAKIQKA-----NPNLRALSETTIYADSRNVEDGRSTQSV 334
DB 486 KEYEKLINEIYDSKFNNNIDLTNFEKMGKRYSYKYEKLTHTPTSPASYENSKNLEKTK 545
QY 335 LMSALDGF--NII-----RYQVTFKNDKGEALDKGNLVTDSKVLTFPKDDEKXT 385
DB 546 ALKYMEDYSLRNIIVKELKYKNLSIKENIELTVE--NIKKDEQLF-----EKIT 598
QY 386 GEDKFNVAIKEDGSMLEFIDTKPVNLSMDKNYNPSPKSNKIYVRNPEFYLRGKISDKGF 445
DB 599 KOENKDEKILEVSDIVKQVOQKVL--MNKIDELKKTQILLKQVNE--LKHNHIVPNVS 653
QY 446 NMELRANESVVDNYLLYGDLDHIDTRDENIKL-----NYK----- 480
DB 654 KOENKOE---PYLLIVLKKKIDKLKVFMPKVESLINEKKNKITQGOQSDNSPSTEGEI 709
QY 481 -----DGDIMDWGMKDYK-----ANGEPDKY 501
DB 710 TQGATTKPGQOAGSALGSDSVQAOAOEQOAPVPVPVPEAKAOVPTPPAPVNNKTENV 769
QY 502 TMDGNGVYLQGTGSDLANA-----KAVGVHYQFLYDVKPEVNIIDPKGNTSIEYADGKSV 556
DB 770 SKLD---YLEKLYQPLNTSYICHKYLIVSHSTNNEKILKQYKLTKEBESGLSSCDPDL 826
QY 557 FNIN-----DKRNGFDG---EIOGHYINGKEYTSFNDIKQIIDKTLNIKIYV 603
DB 827 FNIGNNIPWYSWFSPLNSLSQLFMEIYEKWCNLYLKDNCKIKNLEBKAQVSTVS 866
QY 604 KQPARNTTYKFLINDQGEVSELKPHRYVTIIONGKMSSTIVSEDFILPYKGELEK 663
DB 887 KTLSSSSMQLSLTPQDKPEVANSADTSHSTNNLSKLFPENILS-----LGKNNK 937
QY 664 GYFPDGMELSGFEFGK-----KDAGYVINLSKOTFIKVFKKIEBKEENKPTF--D 713
DB 938 IYQ-----ELIGQSSSENFYEKILKDSPTFNESFTFVSKADINSLNDESRRKLEED 993
QY 714 VSKKDNPOVNHQNLNESHKEDLOREBHQSDSTQDVATVLDKNNISK--STNNP 771
DB 994 INKLKTLQSLFPLNKKYKLEKRLFPDKKTVGKYKQIKGLTYLKEQLSGLNSLNNP 1052

```

RESULT 11

SAZOKI
 Major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (set
 C:Species: Plasmodium falciparum
 C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
 C:Accession: A25120
 R:MacKay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; B
 EMO J. 4, 3823-3829, 1985
 A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falci
 A:Reference number: A91030; MUID:86136024; PMID:3004972
 A:Accession: A25120

A:Molecule type: DNA
 A:Residues: 1-1631 <MAC>
 C:Comment: The merozoite stages of different strains have strain-specific surface antigen
 C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me
 C:Superfamily: major merozoite surface antigen
 C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>
 F:67-84/Region: 3-residue repeats (S-G-T/P)
 F:1614-1631/Domain: membrane anchor #status predicted <MBN>
 F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (asn) (

Query Match 5.4%; Score 218; DB 1; Length 1631;
 Best Local Similarity 18.1%; Pred. No. 0.078;
 Matches 164; Conservative 162; Mismatches 339; Indels 240; Gaps 37;

```

QY 23 KDTTG-VEHHHQBESIK-----EKSFTIDRNISTIRDPENDKL-----LIKKK 69
DB 223 KNVGMEDEYIKKKKTIENINELIESKTTIDKKNATBEEKKLYAOQYDLFLYKQ 282
QY 70 FREVDFTSETGRMEYDYKYDDKNITAYDGDGDLLEYETKLDKIKYIGVLSPSKD 129
DB 283 LBEAHLISVLEKRI-----TLKKNENIK----- 307
QY 130 GHEFIIIGKISNVSKAAKYVYNNYSIEIKATKYDPH-----SKTTFPDIYANINDIV 182
DB 308 ---ELDKINELKPPPPANSNGTPTLTDKNKKIEHEKEIKETIAKTIFKNIDSLFTDPL 364
QY 183 DGLAPAGMRLEFVKNDQKAEIKIRMPKIKETSEYV-----YVSSYGNV-----IELGE 234
DB 365 -----ELEYYLRKN-KNIDISAKVETKESTPEPEYNGVYPLSYNDINNALNELNS 416
QY 235 -GDLKKNKPDNLTKBESGKIYDSEKQOYL--LKONIIL-----RKGYALKTTYP 283
DB 417 FGDLT--NPDYTKPSKNITYDNERKKFINIEIKIKIEKKIESDKSYDRSKSLN- 473
QY 284 GKTDMLEG-----NGVYSKE-----DIAKIQKA-----NPNLRALSETTIYADSRNVEDG 328
DB 474 ---DITKEYEKLINEIYDSKFNNNIDLTNFEKMGKRYSYKYEKLTHTPTSPASYENSKN 530
QY 329 RSTQSVLMSALDGF--NII-----RYQVTFKNDKGEALDKGNLVTDSKVLTFPK 379
DB 531 LERLTVALKYMEDYSLRNIIVKELKYKNLSIKENIELTVE--NIKKDEQLF----- 584
QY 380 DKEVTEGDEKFNVAIKEDGSMLEFIDTKPVNLSMDKNYNPSPKSNKIYVRNPEFYLRGKI 439
DB 585 -EKKITKDNKXDEKILEVSDIVKQVOQKVL--MNKIDELKKTQILLKQVNE--LKHNH 638
QY 440 SDKGFNMELRANESVVDNYLLYGDLDHIDTRDENIKL-----NYK----- 480
DB 639 HVPNSYKQENKOE---PYLLIVLKKKIDKLKVFMPKVESLINEKKNKITQGOQSDNSP 694
QY 481 -----DGDIMDWGMKDYK-----AN 495
DB 695 STEGEITGQATTKPGQOAGSALGSDSVQAOAOEQOAPVPVPVPEAKAOVPTPPAPV 754
QY 496 GFEDKVTMDGNGVYLQGTGSDLANA-----KAVGVHYQFLYDVKPEVNIIDPKGNTSIEYA 550
DB 755 NKTENVSKLD---YLEKLYEPLNTSYICHKYLIVSHSTNNEKILKQYKLTKEBESGLSSC 811
QY 551 DGKSVFNIN-----DKRNGFDG---EIOGHYINGKEYTSFNDIKQIIDKTL 597

```

Db 812 DPLDLFNIQNNI PVMYSMFSLNLSQLFMEIYEKENVCNLYKIKNDKIKNILEAK 871
Qy 538 NIKIVKDFARNTTVAEFLINKDTGSEVSELKPRVVTIIONGKEMSTIVSEEDFILPVY 657
Db 872 KVTSTVKTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFENILS----- 922
Qy 658 KELEKGYOPDGEWISGFEK-----KDAGYVINSKDTPIKPFVKKEIEKKEENK 709
Db 923 LKKNKVIYQ---ELIGQKSSSENFYEKILKSDTFPNESTFNVKSKADINDSLNDESKR 978
Qy 710 PTF--DVSKKKNPOVHNSQLNESHKEDLOREHESQKSDSTKDVATVLDKNNISSK-S 766
Db 979 KULBEDINKLTKTLQLSFPLYNKYKLERLPDKKTVKRYKMOIKKLTLLKEQLSKLN 1038
Qy 767 TTNMP 771
Db 1039 SLNMP 1043
RESULT 12
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
N.Alternate names: gp195 surface antigen
C.Species: Plasmodium falciparum
C.Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C.Accession: S05603; S04850
R.Myler, P.J.
submitted to the EMBL Data Library, April 1989
A.Reference number: S05603
A.Accession: S05603
A.Molecule type: mRNA
A.Residues: 1-1639 <MYL>
A.Cross-references: UNIPROT:P04933; EMBL:X15063; NID:G9896; PIDN:CA33163.1; PID:G9897
R.Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A.Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasm
A.Reference number: S04850; MUID:89345116; PMID:2668887
A.Accession: S04850
A.Molecule type: mRNA
A.Residues: 1504-1639 <MYL2>
A.Cross-references: EMBL:X15063
C.Superfamily: major merozoite surface antigen
C.Keywords: glycoprotein, merozoite, surface antigen
F.1-15/Domain: signal sequence #status predicted <SIG>
F.20-1639/Product: major merozoite surface antigen #status predicted <MAT>
Query Match 5.4%; Score 218; DB 2; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.079;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
Qy 23 KDTTG-VEHHQENESIK-----EKSSTIDRNISTIRDFENDKDKLIRKGFREVD 75
Db 232 KDNVGMEDYIKKNTKIENINLEBSKKTIDKNNATKEBE-----KKKLYQAQY 283
Qy 76 FTSFGKRMEEYDYKDDKGNIAVDGTDLEVEFEKDEISKIYGVLSPKDGHFEL 135
Db 284 DLSTYKQLEB-----AHNLISV-----LR--KRIDTLK-----XNENIKEL 319
Qy 136 GRTSVSKAKVYVYNNYKSIIEIKATKYDFH-----SKWTDFDIYANINDIVOLAVA 188
Db 320 DKINEKPNPPANSNGTPTPLDKNKKIIEHEKEIKELKTKTFNIDSLFTPLP----- 373
Qy 189 GDMRLFVKNDDQKAEIKIRMEPKIKETSEBP-----YVSSYGVN-----IELGE-GDLSK 239
Db 374 -ELEVYLRKKN-KNIDISAKVETKESTEBENYNGVTYPLSTNDINNALNELNSFGDL- 430
Qy 240 NKPDULTKTESGKISDSKQOYL-LKNNIIL-----RKGYALKTTYNPGKTDML 289
Db 431 -NPFYTKSPSKNITYTDERKKFINEIKIEKIKIEKKKIESDKSYEDRSKINL---DIT 485
Qy 290 EG-----NGVSGKE-----DIKIOKA-----NPNLRALSETTYVADSNVDEGRSTOSV 334
Db 466 KEYEKLNEIYISKPRNNIDLTNFEKMGKRSYKVEKLTTHNTPTASYSKNNLEKLT 545

Qy 335 LMSALDGF---NI-----RYQVFTFRQNDKGEAIDKGNLVYDSSKLVLFEGDKDEYT 385
Db 546 ALKYMEDYSLRNITVEKEKELKYKNLISKIENIEITLVE--NIKROEQLF-----EKKIT 598
Qy 386 GADKRFNVAIKEDSGMLTIDTKPVNLSMDKATFPNSKSNKIYVRNPEYELGKISDKGF 445
Db 599 KDNRPDKRILVSDIVAVQVQKVL---MKRIDELKKTQILKNVE--LKNNHVPNSY 653
Qy 446 NWELRVNSVVDNYLIYGDHLIDNTRDFNIKL-----NVK----- 480
Db 654 KOENKQF---PYVILVKEIKEDLKVFMKPVESLINEEKKNIKTBEGSDNSEPTGEHI 709
Qy 481 -----DGDINDWGMKDYK-----ANGPDKV 501
Db 710 TGOATTKRGOQAGSALBEDSVQAQOQKQKQAPVPVPVPEAKAQVPPAPVNNKTENV 769
Qy 502 TDMDNVYLTQGYSLNA-----KAVGHHQGYLYNNVPEVNDIPKNTSIEYADGKSV 556
Db 770 SKLD---YLEKLYERFLNTSYICHKYLIVSHSTMEKILKQYKTYKESSKLSGCDPLDL 826
Qy 557 FNIN-----DKRNNGPDG---EIOEQHYINGKEYTSFNDIKOIIDKTINIKIV 603
Db 827 FNIQNNIPVMYSMFSLNLSQLFMEIYEKEMVCNLYKADNDKIKNLEBAKGVTSV 886
Qy 604 KDFARNTTVAEFLINKDTGSEVSELKPRVVTIIONGKEMSTIVSEEDFILPVYKGELEK 663
Db 887 KTLSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFENILS-----LGKKN 937
Qy 664 GYOPDGEWISGFEK-----KDAGYVINSKDTPIKPFVKKEIEKKEENKPTF--D 713
Db 938 IYQ---ELIGQKSSSENFYEKILKSDTFPNESTFNVKSKADINDSLNDSKRRKLEBD 993
Qy 714 VSKKKNPOVHNSQLNESHKEDLOREHESQKSDSTKDVATVLDKNNISSK-STNMP 771
Db 994 INKLTQTLQSLDLYNKKLKLKLERLPDKKTVKRYKMOIKKLTLLKEQLSKLNSLNNP 1052
RESULT 13
G90603
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C.Species: Mycoplasma pulmonis
C.Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C.Accession: G90603
R.Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A.Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A.Reference number: A99512; MUID:21267165; PMID:11353084
A.Accession: G90603
A.Status: Preliminary
A.Molecule type: DNA
A.Residues: 1-1546 <KTR>
A.Cross-references: UNIPROT:Q98P17; GB:AL445566; PID:G14090150; PIDN:CAC13908.1; GSPDB:G
A.Experimental source: strain UAB CTIP
A.Genetic code: GCG3
Query Match 5.4%; Score 217.5; DB 2; Length 1546;
Best Local Similarity 20.2%; Pred. No. 0.077;
Matches 180; Conservative 125; Mismatches 336; Indels 249; Gaps 40;
Qy 1 KLGELAESKRFNLNGKSGSLK---KDTTVEHHQENESIKESSTIDRNISTIRD 56
Db 103 KSKELAESQVGVSVNNEKNRILAEKATKNTKONPSAKSQNLDTSGDHLNKSQKSNIN 162
Qy 57 FENKDLKLIKKEPREVDFTSETGKRMEEYDYKD-DKGNIAVDGTDLEVEFEKDE 115
Db 163 IKNEQSKTLQTLNNEEHSANLQTONLILESNTQODLDKPN-----NSNKEKEKLN 214
Qy 116 IKSRTYGVLSPSKQGHFILLK-ISNVSKNKKVYVYNNYKSIIEIKATKYDFFSKMTPTDL 174
Db 215 VLS--FNKQEAIVKDSFEFFYFREIVSKLSESKINIDKRVDEITLDTKTRADF-----SFEQ 267

```

QY 175 YANINDIVDGLAFAGMRLFKVNDQKKAIEKIRMEKIKETKSEYVYSSYGNVIELGE 234
DB 268 Y--IEKIKOLFVKWAST-----IKDTYQ-----TNKIFLTNDEY----- 298
QY 235 GDLNRKPNLTQME-----SGKIYS-----DSEKOYLLKDNITLIRKGYAL 276
DB 299 --IKKNESNLKDKKPFNNFENIEIYKFLKDLALVEIQALHKOYLEKKEIITKGEIYD 356
QY 277 KATTYNGKTDMLEGNG-VYSKEDIKTI-----OKANPNLRALS 314
DB 357 KAAAFIKSELKSKAGLFLSKQSAKITOLINHIISRPPEAPENLELSKAMGLINEL 416
QY 315 ETTIYADSRNVEDGRST-QSVLSALDGFNIIRYOVFTPKMDKGEALIDKGNLVYDSSK 373
DB 417 KPEIYVEKEGVEIAYPTLQDAISNAQDGKIFL-----NKMLKDKR--SIYVDKN- 464
QY 374 LVLFQDDKEYTEED--KRVNEAIKEDGSMFLDTYKPVNLSMDKATFNPS----- 421
DB 465 IITFAKSNVITITKDKSSKFTMFIYQKALTFEIAEPSSQSINLNGIGTSFKDESSLVKI 524
QY 422 -KSNKIYVRNPEFYLRGKISDKGFMWELRVNESVVDNYLIYGDLDHIDTRDFN----- 474
DB 525 EKNAKLVAKTGTAFINSKFSKYG-----SVFEN--YGSVIEGAKIMNNVNSBSG 572
QY 475 -----IKLVKDDIDMDGKDYKANGFPDK-VTDMGNYVLTQYSDLNKAKAVGYH 525
DB 573 GIIRNVGSLGTFPKNBEI-----RDNISITG--DKGIISQGNIAISGGSIDGKKS----- 620
QY 526 YQFLYNNVKEPVNIDPKNTSIEYADGKSVENINDR-----NNGPDGEIQECHIYING- 580
DB 621 PFRSLINLR-KTMINNPGSGIVNNAASVKSLFEIDNSKIQISNNALINPGSSAIFLKN 679
QY 581 -----KEYTSFNDIKOIIDKTLINKIV----- 602
DB 680 STMHLAGSLKIKKESBQRIEVLVLPLOAKLISPKNITSLDNYQKLSAIFKIPSVKN 739
QY 603 VQDFAR-----NTTYKEFLINKDT-----GEVSELKPHRVVTIIONGKEMSTIVSEED-DF 652
DB 740 INDFKIVPLVWNTKKEFFKLPMPDTKLFVNFYKTLKQNHDLIQSGPFESTIEKIKIDELDF 799
QY 653 -----IIPVYKGELEKGYQF--DGEWISGFEGRK-----DAGYVYNLSK 669
DB 800 YFRPTAAVKKLILITQLVRTIPKTHKFWMAFEYENPFLQKWEYETINELIRIDPFYLDIAY 859
QY 690 DPFIRKVFFKIEKKEENKPTFDVSKKONPQVNSQLNESHKEDLQR 739
DB 860 PEFV-----BNGKMLPKPEY-IHTNVNVPVLEH-----PRNEDVAR 894

```

RESULT 14

B42771
 reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
 C1Species: Plasmodium vivax
 C1Date: 12-Mar-1993 #sequence_rev1sion 12-Mar-1993 #text_change 09-Sep-1997
 C1Accession: B42771
 R1Galntki, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
 Cell 69, 1213-1226, 1992
 A1Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
 A1Reference number: A42771; PMID:92315338; PMID:1617731
 A1Accession: B42771
 A1Status: preliminary; nucleic acid sequence not shown; translation not shown
 A1Molecule type: DNA
 A1Residues: 1-1252 <GAL>
 A1Cross-references: GB:M88098, NID:g160627, PID:g160628
 A1Experimental source: strain Belem, merozoites
 C1Genetics:
 A1Gene: RBP2

Query Match 5.4%; Score 215.5; DB 2; Length 1252;
 Best Local Similarity 20.9%; Pred. No. 0.071;
 Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;

```

QY 2 LGEIAESKFK---NLNGKESGLKDTGTGVEHHQDNEESIKESFT- IDRNI----- 51
DB 266 LSEIKKDKCTTEISNSRG--KDKIEFLKFKPNESENKNNKININENINRSSEYQ 322
QY 52 -----STINDFEKDKLIKLR--PREVDFFS 78
DB 323 LKDIEDAEKQASTKVELFPHHBTTSNIPKSEBILIGVEFKSOCKINKADIMEKIERHNS 382
QY 79 ETKRMEEDYKCKDDKNIIAYDDGTDELEYETEKDEISKITYGVLSPKDGHFELIGKI 138
DB 383 ELQYQKGFQENLANKLNPENTDMAED-ELANDKSNKAVLIETNLESYKH-----LSEI 437
QY 139 SNVSKNAKYVYANNYKSI--EIKATKYDFHSKTM-----TFDLVYAN-INDIVDGLAFACD 190
DB 438 TYNKQGEKITY-SKADIMQIKATSENPAETKLEKVKDQSNVYVYLNQITF-----E 490
QY 191 MRLFYDND-----QKAEIKIRMEKIKR-----TK-----S 218
DB 491 RMLIVTEKRLNGIDSTTINIGALKESKNGYEIGFLEKLEIIGNKRLKVDITYKKSINS 550
QY 219 EYPVYSSYGNVIELEGGDLKRNKPDNLTMESGKIYSDSEKOQYLLKDNITLIRKGYALV 278
DB 551 TYGNFSSLPNNPDLDNDYDPNKNINDYENKM--GEIYNEEGSLNKISENL--RNASENT 605
QY 279 TTYNPEKTDMLEGNGVYSKEDI--AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVL 335
DB 606 SDYNSAKTLRLBA-----QKEKYNLKNKEEAKHYLR-----DYKKVESPR----- 646
QY 336 MSAIDGFNIIRYOVFTPKMDKGEALIDKGNLVYDSSKLVYFGKDDKEYTGEDKRVNAEI 395
DB 647 -----FIFNNK--ESLDKINEMIKKEQLTV-----NEGHGVNKKQLEVENI 683
QY 396 KEDGSMFLDTKPVNLS-----MDKN-----YENPKSNKIYVRNPEFYLRGI 439
DB 684 KE-----LVDEN--NLSDLKQATGKNBEIQKITHSTLKNKAKTTLIGHDTSAKTYGKI 736
QY 440 SDKGFNMELRVNESVVD-----NYLIYGDLDHIDTRDFNLIKLVND----- 481
DB 737 TP-----ELALTELLGDAKAKTQAQELKPKSKANNVVLLETENNSKQNTNELDVHKNIQDAYK 790
QY 482 -----GDINDGKDYKANGFPDKVTDMGNYVL-----QTSYSLNMAKAVGYH 525
DB 791 VALEILAHSDDEIDTKQD-----SSKLIEMGNQIYLVKVVLIQYKKNKISSIKSEEAVS 844
QY 526 YQFLYDNYK-----PEVNIIDPKNTSIEYADGKSVENI-----NDKKNNGPDGEIQO 574
DB 845 VKI--GNVSKKSELSKITCSDPSYDNIITALEKQTELQNLNNSFTQEKNTNNSDKLR-- 900
QY 575 HIYNGKEYTSFNDIKOIIDKTLINKI-----VVND----- 605
DB 901 -----KIKTDFESLKNML-KTLBGEVNAKLKASDNHHEVQSKSEPVNPALSEIEKEED 953
QY 606 -PARNTTYKEFLINKDTGEVSELKPHRVVT--IONGKEMSTIVSEEDFILPVYKGELE 662
DB 954 ILSLWATLDELKAKRTCEVSHYKLIKQIVTKEISDDTELINTIENK--YKAYIAYIK 1009
QY 663 KAYQDGMWISGFEGRKQAGVYNLSKQTFKPV-----PKIIEKKEENKPTFD-- 713
DB 1010 KNYE-----DTVQD--VLTNHEPNTQVSNHEPTNDKNSKSEELTKAVTDSK 1057
QY 714 --VSKKK-----DNPQVN-----HSQNLNSHR-KEDLOREBHSQ 744
DB 1058 TIISLKGVIIIVNNTENTNTESSAKETALYNELKKNKTSLNEIYQISNVEVKAQEMKS 1117
QY 745 KSDSTKDVTA--TYLD--KNNISKSTTN 770
DB 1118 NADKIVDSKIPNTVLDQKSNIVTNQSHINN 1149

```

RESULT 15

A45597
 DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmodi
 C1Species: Plasmodium falciparum

C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #ext_change 09-Jun-2000
C:Accession: A45597
R:Li, W.B.; Bzik, D.J.; Tanaka, M.; Gu, H.M.; Fox, B.A.; Inselburg, J.
Mol. Biochem. Parasitol. 46, 229-239, 1991
A:Title: Characterization of the gene encoding the largest subunit of Plasmodium falciparum
A:Reference number: A45597; M01D:92018020; PMID:1656254
A:Accession: A45597
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2339 <L11>
A:Note: sequence extracted from NCBI backbone (NCBI:61099, NCBI:61101)
C:Superfamily: Plasmodium DNA-directed RNA polymerase II large chain
C:Keywords: nucleotidyltransferase; transcription

Query Match 5.3%; Score 214; DB 2; Length 2339;
Best Local Similarity 21.6%; Pred. No. 0.18;
Matches 191; Conservative 134; Mismatches 306; Indels 254; Gaps 50;

QY 24 DTTGVHHHQB--NEESIEKSSFTIDRNIISTIRDF--ENKDLKLIKKKPREVDPTS 78
DB 1292 DNTYVEQIEMKEISKNKTEKQSPK-----GTIRDMHDESEQMNKFTKKAK---FPI 1342
QY 79 ETGKMEEDYKDDKGNIIAYDDGTDLEYTEKDEIKSYI-----G 122
DB 1343 EK-KKGKMECNDDELYNNTQYD--NIQYNNISCNVYKSQMLENTHQVNNDLSEFIKN 1398
QY 123 VLSPSKQCH-----FELIGKISNVSNAKAYVGNMYKSIKATKYDPHAKTWTPTDL 174
DB 1399 VILPEKYSIHFVNDYDANVVEIKMLMDKKIIFLNSSEKN--VQSKYNRMSKNLAKKI 1456
QY 175 YANINDIVDG-----LAPAGDMRLFVKNDQ--KKAELIKIMPEK---IKETKSE 219
DB 1457 EI-INNIYENKEKKILRMWTKMDNDNDYSSDDSIIAKKIIKIKNKEKRYHPKEKEN 1515
QY 220 YPVSSYGVNIIELGEGDLSKNKPDNLTK-MESGKIYSDEKQOYLKXNIIIRKGYALKY 278
DB 1516 PDR-NNYKMITDNN 1569
QY 279 TTYNPGKTDML--EGNGVYSKEDIKIQANPIL-----RALSFTTIYADS-- 322
DB 1570 TMYN---TWIYRNDGCIYEKETNNNELTTNSMCDKNNDFSEFFNNINENDLYDNKY 1626
QY 323 ----RNVEDGRSTQSVLSALDGFNIIRYQVF--TFPMNDKGEALDKGNLVTDSSKL 374
DB 1627 YRQIFKNVIGFVSFEYVESYQHYILFPYEIILKMTSFLLEYLTETIIPNIFLHTQLSK- 1685
QY 375 VLFQGD---DKXYTGEDKFNVEAIKEDGSMFLPDTKPVNLSMDKNYFNPSKNTIYVRNP 431
DB 1686 ---KEKPTHQKNTGKMKIYIEIKK--WLF--KAINI--YKYFSFKKSIEL----- 1728
QY 432 EFLYLRKGISDKGFNNELRVNESVVDNVLVYGLDHIDNTRD-----FNI-----KLVNKO 481
DB 1729 ----IKKKOYFNYYIIR-NDYISHRYYI--DYSFINLQOLYLPFPIFYKPKYISTP 1779
QY 482 GDIMDKGMKDYKANGFPDKVYTDMDGNVYLQTYGSDINAKAVGVHYQFLYDNVKEVNIDP 541
DB 1780 GD---AVGSIASQSIGEPQTQTLKTFHFAGVASMNV-TLGV-----PRIKEIINASN 1828
QY 542 KGNSTIEVADGKSVFNINDKXNGF-----DGEIOGHYIYNGKEYTS-----F 586
DB 1829 SIQTFI-----LNTPELVNDNVPALMMSKLEKTTIRIDICMYIK-EDYTSRGVFLSVKF 1882
QY 587 ND-----IKOIIDTKLNI-----KI---VVKDFARNTTVK--EPIINKDT 621
DB 1883 NEELLQKLPLNANVNIKOIILKQSHINKIKINKIHINVINKYKHLISLKNDEFIR----- 1938
QY 622 GEVSELKPHRYVTYTIONGKEMSTIVSEDFIIPYKGELEKGYQFPGWEISGFEGK--K 679
DB 1939 FQWESLKKGLLDLLIYGDKIDKRCIIEKEDI-----EYTDNDEIDCD 1980
QY 680 DAGYVINLSKDTFIKVPFKKIEKK--BEENKPTFDVSKKDNPOVNSQLNESHKED 736
DB 1981 DMDEYVNVSGT-----ELYERKCSKEENKNAIRVKEEID-----DN 2019

QY 737 LQREE-----HSQKSDSTQDVTATVLDKNNISSKSTYTN 770
DB 2020 LEKENIITYVSEKDSVNLKSEKKKDIN-----DNNNNNDNNNNNN 2060

Search completed: February 16, 2005, 19:21:34
Job time : 51 secs